THE GENETIC AND MOLECULAR ANALYSIS OF PRIMARY CILIARY DYSKINESIA

Victoria Helen Castleman

A thesis submitted for the degree of Doctor of Philosophy in the University of London

May 2008

General and Adolescent Paediatric Unit
Institute of Child Health
University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, Victoria H. Castleman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Primary Ciliary Dyskinesia (PCD) is a recessively inherited disorder caused by cilia and sperm flagella dysmotility associated with axoneme ultrastructural defects. Symptoms include recurrent respiratory tract infections, sinusitis, bronchiectasis, subfertility, and laterality defects due to defective embryonic nodal cilia. PCD is genetically heterogeneous and two genes, *DNAI1* and *DNAH5,* account for 38% of cases.

To identify new PCD genes, genome wide linkage screens were undertaken in consanguineous PCD families using homozygosity mapping: (1) five Pakistani families with missing inner and outer dynein arms; (2) two Arabic families with central pair agenesis and no dextrocardia.

Three disease loci were mapped on chromosome 11q23.3-24.3 and 17q21.31-22 (Pakistani), and chromosome 6p21.31-21.1 (Arabic), with peak multipoint LOD scores of 3.6, 6.0 and 6.7, respectively.

Comparative bioinformatic analysis identified 7 positional candidate genes which were subjected to mutational analysis. A 3 bp deletion in *C6ORF206* at 6p21.31-21.1 was revealed in affected Arabic PCD individuals, resulting in a predicted in-frame loss of a C-terminal lysine residue, K268. The protein encoded by *C6ORF206* was identified as homologous to the *Chlamydomonas reinhardtii* radial spoke head protein, RSP9.

Functional work was undertaken to determine if the K268del mutation was pathogenic. *RSP9* is mutated in the paralysed flagella *Chlamydomonas* mutant, *pf17,* and transformation of *pf17* with wild-type *RSP9* rescued motility. However transformation of *pf17* with *RSP9* R261del (equivalent to human K268del) rescued motility to a lower level, resulting in an ineffective swimming stroke.

Expression of *RSP9* was investigated by *in situ* hybridisation, and zebrafish RSP9 knock-down morphants were created to investigate its role in vertebrate ciliary function. Although expressed at the vertebrate embryonic node, knockdown of RSP9 function did not affect laterality in zebrafish, however it did cause dysfunction of the nasal cilia.

This data suggests that C6ORF206/RSP9 functions as a ciliary protein in ciliated organisms and that the K268del mutation likely causes PCD without *situs inversus.*

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A	adenine
ABI	Applied Biosystems (Perkin Elmer)
AD	autosomal dominant
AKAP	A-kinase anchoring protein
APS	ammonium persulphate
AR	autosomal recessive
BLAST	Basic Local Alignment Sequence Tool
bp	base pair
C	cytosine
CBF	ciliary beat frequency
cDNA	complementary DNA
cM	centimorgans
CP	central pair
DHC	dynein heavy chain
DI	Dysmotility Index
DIC	dynein intermediate chain
DLC	dynein light chain
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DRC	dynein regulatory complex
EDTA	ethylene-diamine-tetra-acetic acid
EM	electron microscopy
EST	expressed sequence tag
EtOH	ethanol
FAM	5'-fluorescein phosphoramidite (ABI fluorescent label)
G	guanine
g	gram(s)
h	hour(s)
HBD	homozygosity by descent
HBS	homozygosity by state
HEX	5'-Hexachlorofluorescein phophoramidite (ABI fluorescent label)
HLOD	heterogeneity logarithm of the odds

Hz	hertz
IDA	inner dynein arm
IFT	intraflagellar transport
II	Immotility Index
kb	kilobase
kDa	kilo Daltons
L	litre(s)
LD	linkage disequilibrium
LOD	logarithm of the odds
LRA	left-right axis
M	molar
mA	milli-amps
MAPs	microtubule associated proteins
Mb	megabase
µg	microgram
mg	milligram
$MgCl_2$	magnesium chloride
min	minute(s)
µl	microlitre
ml	millilitre
mM	milliMolar
mRNA	messenger RNA
MT	microtubule
MTOC	microtubule organising centre
NaCl	sodium chloride
NCBI	National Centre for Biotechnology Information
nm	nanometres
OD	optical density
ODA	outer dynein arm
ODA-DC	outer dynein arm docking complex
OID	outer-inner dynein linker
PCD	primary ciliary dyskinesia
PCR	polymerase chain reaction
pmol	picomole

RNA	ribonucleic acid
RS	radial spoke
RT-PCR	reverse transcriptase PCR
s/secs	seconds
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
T	thymine
TAP	Tris-acetate-phosphate
T_m	melting temperature
Taq	*Thermus aquaticus* DNA polymerase
TBE	tris-borate EDTA
TE	tris EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TET	5'-tetrachloro flourescein phosphoramidite (ABI fluorescent label)
Tris	2-amino-2-(hydromethyl) propane- 1, 3 diol
UTR	untranslated region
UV	ultraviolet
V	Volt

ACKNOWLEDGEMENTS

Firstly, I would particularly like to thank my supervisor Dr. Hannah Mitchison for her vital help and advice during the course of this project and for all her valuable comments and advice on my thesis.

I would also like to thank Keith Parker for his practical assistance and his mints, Dr. Eddie Chung for his advice with the linkage analysis and intellectual input and Professor Mark Gardiner for scientific guidance. I would like to thank the Medical Research Council for allowing me to pursue this project.

I would like to thank Dr. Saul Purton, Chloe McCann, Dr. Leila Romio and Dr. Robert Hirst for their practical assistance and advice in their specialised areas.

I would like to thank my friends and colleagues in the General and Adolescent Paediatrics Unit, especially Barry, Kate, Jacqui and Christina, not only for their help around the lab and on the many technical computer issues, but also for their calming input. I would also like to thank Kathy and Ellie for the entertaining coffee breaks, as well as Emma, Fairy, Lorna and Laura for always being on the end of the phone during my fits of pique. Thank you all for holding me up during the dark days and giving me a place to stay when I was homeless.

I would like to thank Louis, my lovely boyfriend and now my sweet angel in heaven, for being in my life, even though you were taken so suddenly. Thank you for being somewhere to escape when the going got tough and for giving me something to look forward to at the end of hard days in the lab, even though you never really understood exactly what it was I was doing!

Finally, a big thank you to Matt for caring and letting me use your flat as a quiet haven to write up in and my family: Mama, Git and Tommy C for their inspiration, kindness and support which often went far beyond the call of duty.

For my Dad

1 CHAPTER ONE

Not waving but drowning

"Nobody heard him, the dead man,
But still he lay moaning:
I was much further out than you thought
And not waving but drowning.

Poor chap, he always loved larking
And now he's dead
It must have been too cold for him his heart gave way,
They said.

Oh, no no no, it was too cold always
(Still the dead one lay moaning)
I was much too far out all my life
And not waving but drowning."

Stevie Smith, 1957

1.1 *Introduction*

Primary Ciliary Dyskinesia (PCD) (OMIM 242650) is a genetically inherited disease characterised by abnormalities in ciliary ultrastructure and function and consequent reduced mucociliary clearance (Bush et al., 2007). PCD affects the upper and lower respiratory tracts and results in recurrent/persistent respiratory infections, sinusitis and otitis media (van's Gravesande and Omran, 2005). Male subfertility is also present as a result of immotile sperm flagella (Geremek and Witt, 2004), and randomised left-right axis determination (usually *situs inversus* – complete mirror image reversal of internal organs) which is presumed to be due to ineffective embryonic nodal cilia (Essner et al., 2002). Usually clinical manifestations begin in the first year of life and respiratory problems often progress to permanent lung damage (bronchiectasis) (Coren et al., 2002).

This study involves the molecular genetic investigation of PCD. First, ciliary structure and function is described then the clinical features, diagnostic and genetic aspects of PCD with a summary of the current status of genetic research and the technical advances that have had an impact upon disease gene identification.

1.1.1 The role of cilia in the human body

Cilia are hair-like structures 5-7 μm in length (Afzelius et al., 1985) extending from the surface of epithelial cells (Figure 1.1). Cilia evolved from flagella, which appeared early in evolution to provide unicellular organisms with motility in water (Ibanez-Tallon et al., 2003). Cilia are structurally related to flagella and are present on most cells of the body, including respiratory epithelial cells, sinuses and middle ear, brain ependyma and female fallopian tubes (Pan et al., 2005). Cilia perform a diverse array of biological functions including beating movement to move fluid across the surface of cells and a sensory role to detect and assess the surrounding environment of the cell. Respiratory epithelial cilia are important in mucociliary clearance and move mucus up towards the throat for swallowing. Motile cilia lining the ependymal epithelium beat to move cerebrospinal fluid. Cilia lining the fallopian tubes act to propel oocytes towards the womb for fertilisation. Cilia in the efferent ductules of the testes propel immature sperm to the vas deferens, where they mature and become motile upon growth of a single motile flagellum. Monocilia are present transiently in the embryonic node and are involved in creating the nodal flow that breaks the left-right asymmetry in laterality determination (Nonaka et al., 1998).

Figure 1.1: Respiratory epithelial cilia
Cilia are found on most cells of the body, including those lining the trachea as shown by electron microscopy in this figure. Goblet cells are also present in the trachea and secrete mucus, which is then propelled up into the throat by the whip-like movement of cilia and is then swallowed.

1.1.2 Classification of cilia types

There are at least eight categories of cilia in the human body. These all have an invariant basic pattern – nine microtubule doublets in a ring (described in more detail in section 1.1.4), but differ in several details including length, beat pattern and structure including presence or absence of central microtubules, motor dynein proteins and associated structural components (Table 1.1). Motile cilia and flagella have been retained and employed for a number of biological functions, including left-right axis determination, cerebrospinal fluid flow, mucociliary clearance (Ibanez-Tallon et al., 2003) and spermatozoa motility (Sapiro et al., 2002). Non-motile or sensory cilia function in various sensory roles including retinal, auditory and renal physiology, for example olfactory cilia present in odour receptors and cilia present in the kidney (Salisbury, 2004). Modified cilia are found in retinal rod cells and form a "connecting cilium" between the inner and outer segments (Arendt et al., 2004).

Malfunctioning of any one of the different types of cilia can have consequences on normal physiology (Table 1.1). The most prominent abnormality involving motile cilia is PCD and the clinical phenotype of PCD patients correlates with the tissues and parts of the body in which their dysmotile cilia are found.

Table 1.1: Categories of cilia in the human body. Adapted from Afzelius, 2004. 9+2 refers to axoneme structure of 9 peripheral microtubules and a central pair. 9+0 refers to axoneme structure of 9 peripheral microtubules and no central pair.

1.1.3 Introduction to *Chlamydomonas reinhardtii* as a PCD model organism

Chlamydomonas reinhardtii (Figure 1.2) is a eukaryotic biflagellate unicellular alga, which propels itself through water using its two flagella (Lefebvre and Silflow, 1999). These flagella have a similar ultrastructure and high conservation of protein homology to human respiratory cilia and sperm flagella. *Chlamydomonas* is therefore a key organism that has historically contributed extensively to the understanding of human cilia structure and function. Much current knowledge of axonemal protein composition and organisation arises from studies in *Chlamydomonas*. They have been used extensively in genetic and biochemical analysis as a model for the investigation of cilia function and assembly (Pazour et al., 2005), aided by the analysis of the *Chlamydomonas* genome, which has advanced understanding of this eukaryotic cell and revealed genes associated with flagella function (Merchant et al., 2007). This model is advantageous because it has a haploid genome so that genetic mutations are fully expressed and are not masked by wild type genes.

Mutants with dysmotile or paralysed flagella are fully viable and are easily detected by their reduced motility. The flagella proteome has been completed and showed motor and signal transduction proteins with homologues associated with human disease, such as cystic kidney disease, male sterility and hydrocephalus (Pazour et al., 2005). Furthermore, some unexplained components were found to be part of the *Chlamydomonas* proteome, suggesting that the complexity of the axoneme is greater than originally thought. Many *Chlamydomonas* mutants which demonstrate a lack of flagella movement have been screened for mutations and genetic loci and from this a list of candidate genes and chromosomal locations have been suggested for the mapping of genes that cause PCD (Geremek and Witt, 2004).

Figure 1.2: Biflagellate algae *Chlamydomonas reinhardtii*. Image adapted from www.botany.hawaii.edu/BOT201/Algae/Bot%

1.1.4 The structure of cilia

The eukaryotic cilium is a dynamic structure consisting of over 250 proteins (Dutcher, 1995; Pazour et al., 2005) that is variable in length depending upon location in the body (Table 1.1). The protein composition of the core ultrastructure of cilia/flagella, known as the axoneme, has been well studied and also visualised by electron microscopy (EM) and more recently by electron tomography (Nicastro et al., 2006). The motile ciliary axoneme consists of 9 peripheral outer doublet microtubules that surround a central pair (Figure 1.3A). This is known as the 9+2 structure. Each peripheral doublet is constructed from an A and a B tubule, composed of 13 and 11 tubulin protofilaments, respectively. These protofilaments consist of heterodimers of α and β tubulin, which are the most abundant proteins (70 % total protein mass) in the cilium (Cosson, 1996). Each outer doublet is attached to two rows of dynein arms (inner and outer), discussed in Section 1.1.4.2.1. Radial spokes create a link between the peripheral and central pair of microtubules and are thought to play both a structural and signalling role in ciliary beating (Yang et al., 2006), discussed further in Section 1.1.4.2.2. The dynein regulatory complex (DRC) is thought to coordinate the activity of the dyneins and is located at the base of the radial spokes in close association with the inner dynein arms (Gardner et al., 1994). Nexin links link the peripheral microtubules and restrict the extent of the sliding (Cosson, 1996). The underlying ciliary ultrastructure has a direct influence on the whip-like waveform of cilia that is effective for mucociliary clearance and disruption of this structure affects the motility and therefore the function of these structures (Ibanez-Tallon et al., 2003).

Figure 1.3: Diagrammatic representation of the structure of the flagella/ciliary axoneme and transmission electron microscopy of a cilium in cross-section.
A) Ciliary axoneme - diagram obtained from Gartner and Hiatt, *Colour atlas of histology,3rd ed.* Lippincott Williams & Wilkins, 2000. Nine outer doublet microtubules formed of an A-microtubule and a B-microtubule arranged around a central pair. Inner and outer dynein arms are directed towards the next microtubule pair. Radial spokes connect the central pair and peripheral microtubules. Nexin links provide structural links between outer microtubules. B) EM of cilium in cross section, obtained from Google Images, reference www.sun.melanschool.org. C) Longitudinal view of a peripheral pair of microtubules with associated proteins – taken from El Zein *et al*, TRENDS in Genetics, 2004. Outer dynein arms (red) spaced every 24 nm (short arrow), inner dynein arms form different complexes (pink, grey, blue and yellow) with a 96 nm repeat (long arrow), radial spokes (brown) are located along the axoneme with 96 nm periodicities. Dynein regulatory complex (orange) is involved in regulation of IDAs via radial spoke interactions.

1.1.4.1 Microtubules

Microtubules are cylindrical structures made up of alternating heterodimers of α- and β- tubulin (Mandelkow and Mandelkow, 1995) (Figure 1.3). Microtubules exist in two forms, those found in the ciliary axoneme and those found in the mitotic spindle of dividing cells (Brinkley, 1997). They are dynamic structures that grow from microtubule organising centres (MTOCs) such as basal bodies and centrosomes (Wade et al., 1998) (Figure 1.3). Cilia and flagella formation begins when the basal body docks to the plasma membrane and serves as a template for axoneme assembly. The axoneme and the surrounding ciliary membrane, which is continuous with the plasma membrane, projects from the cell body in a process known as compartmentalised ciliogenesis accomplished by intraflagellar transport (IFT) (Section 1.1.6) (Scholey, 2008). The cilium therefore elongates and creates a separate compartment that is separated from the cytoplasm by transition fibres. By contrast, in cytosolic ciliogenesis, (e.g. in *Drosophila* sperm cells) the entire axoneme is assembled inside the cell and is later extruded (Baker et al., 2004). Microtubules have an inherent polarity, whereby the "plus end" grows faster than the "minus end" and in the case of axonemal microtubules, the plus end is always located at the tip of the cilium.

Tubulin is the major component of microtubules and is a highly conserved protein of approximately 450 amino acids. As mentioned previously, it exists in 2 major isoforms in the microtubules of the ciliary axoneme, α and β. Tubulin protein undergoes post-translational modifications prior to integration into the ciliary axoneme, however knowledge on this is as yet limited.

1.1.4.1.1 The axoneme central microtubule pair

Regulation of motile 9 + 2 cilia and flagella depend on interactions between the radial spokes and the central pair apparatus (Smith and Lefebvre, 1997; Mitchell, 2004). Studies of the axoneme central pair microtubules have used a wide range of species including flagella isolated from *Chlamydomonas reinhardtii* (Goodenough and Heuser, 1985). The central pair are orientated with respect to the central pair of

adjacent cilia and this determines the direction of the ciliary beat (Mitchell, 2004) (Smith and Yang, 2004). The central pair of microtubules (C1 and C2) (Figure 1.4) are structurally and biochemically asymmetric. Each individual microtubule of the central pair are made up of 13 tubulin protofilaments and have a distinct protein composition from that of the outer doublets; they are more like cytoplasmic than axonemal microtubules in their lability, in that they more readily undergo change or breakdown (Dutcher, 1995). Thin section EM of *Chlamydomonas* flagella revealed that the central pair are surrounded by specific protein projections that appear to wrap around the central microtubules (Figure 1.4). The longer projections around the C1 microtubule repeat along the axoneme with 32 nm repeat periodicities, whereas the smaller C2 projections repeat every 16 nm (Mitchell, 2004).

Comparison of the structural relationships within central pair complexes from wild type and central pair mutant *Chlamydomonas* strains has provided a model of radial spoke interaction with sites on the central pair (Mitchell, 2004). During the flagella bend, the central pair twists so that the C1 microtubule is closest to the doublets on the outside of the bend. This allows the central pair projections to interact with the radial spokes attached to the active dyneins, while the other central pair projections interact with the radial spokes on doublets with inactive dyneins. It has therefore been suggested that specific projections associated with the C1 and C2 microtubules contribute to regulating motility (Wargo et al., 2005) (Lechtreck and Witman, 2007) by regulating dynein through interactions with radial spokes and microtubule associated proteins.

Figure 1.4: Diagrammatic representation of the *Chlamydomonas* flagellar central pair.
Taken from Mitchell, 2003 (Mitchell, 2003) Central pair apparatus viewed longitudinally. Proteins associated with C1 microtubule are 1a, 1b, 1c and 1d. Proteins associated with C2 microtubule are 2a, 2b and 2c. sh indicates protein sheath. 1a, 1b, 2a and 2b projections (pink and blue) run perpendicular to the central pair axis at the radial spoke interaction surface. Green colour indicates microtubule bridges that connect the two microtubules.

1.1.4.2 Microtubule associated proteins (MAPs)

Microtubule associated proteins (MAPs) is the term given to other structural components of cilia. They include proteins that are involved in the maintenance of axonemal structure during the bending of the cilia, such as nexin links that connect adjacent outer doublet microtubules (Figure 1.3) and tektin filaments that are found at the junction between the A and B tubules of the outer microtubule doublets (Pirner and Linck, 1994). Radial spokes are MAPs connecting the peripheral and central microtubules and are discussed in detail is Section 1.1.4.2.2.

The dynein arms are the best characterised of all the MAPs and are divided into two types: inner and outer arm dyneins. They are large molecular motors that were first isolated from *Tetrahymena* cilia (Asai and Wilkes, 2004). They provide the cilium with the force for propulsion of its coordinated bending (Asai and Wilkes, 2004).

1.1.4.2.1 *MAPs: Dyneins*

Axonemal dyneins repeats have been extensively studied in *Chlamydomonas*, accounting for 15 % of the total protein mass of the *Chlamydomonas* flagellum (Porter and Johnson, 1989) (Witman and Minervini, 1982). Dyneins are multisubunit ATPase motor protein complexes that occur periodically along the length of each of the nine peripheral A microtubules, in two rows (as inner and outer dynein arms) (Woolley, 2000). The outer and inner dynein arms have been shown to be functionally distinct as well as of different composition and they play separate roles in the generation of the ciliary beat. Outer dynein arm docking complexes (ODA-DCs) attach outer dynein arms to the peripheral A-microtubules and repeat with 24 nm periodicity along the length of microtubules (Takada et al., 2002). Inner dynein arms attach to microtubules in doublet or triplet groups every 96 nm (Smith and Sale, 1992) by specialised adaptor molecules that designate the precise location of each inner arm subform (Smith and Sale, 1992). The dynein arms function by reversibly binding to the adjacent B-microtubule and using ATP hydrolysis to provide the energy for sliding along microtubule filaments (Wemmer and Marshall, 2004). This sliding is converted into ciliary bending by the constraints of connecting proteins such as the radial spokes, although the exact mechanism that produces the ciliary waveform is not yet entirely understood (Woolley, 2000).

The *Chlamydomonas* outer dynein arms (Figure 1.5) are large complexes that contain three dynein heavy chains (HCs) of molecular mass >500 kDa (α, β and γ), two or more intermediate chains (IC1 and IC2) and four to eight light chains (LCs). The N-terminals of HCs bind to ICs and LCs. The globular C-terminals of HCs consist of six AAA (ATPases Associated diverse cellular Activities) domains arranged in a hexameric ring with a short extension that binds to microtubules in an ATP-sensitive manner. This globular head contains the first P-loop, which provides the nucleotide binding site where ATP binds. Four other P-loops exist in dynein HCs, although the function of these is not yet known. It has been suggested that they allow regulation of the dynein molecule. There are six IC isoforms that have specific functions in the dynein complex, functioning to regulate HC activity and allow the attachment of the dynein arm to microtubules (Witman, 1992) (Lo et al., 2006). LCs are thought to function in both outer arm assembly and motor function (DiBella et al., 2005).

Analysis of human homologues of proteins associated with *Chlamydomonas* outer arm, revealed that a total of 24 human genes are thought to encode outer arm subunits and/or components necessary for assembly and that 12 out of 14 known *Chlamydomonas* outer arm subunits have at least one likely orthologue in humans (Pazour et al., 2006).

Figure 1.5: ***Chlamydomonas*** **flagellar Outer Dynein Arm (ODA)**
The α, β, and γ heavy chains of the outer arms are labelled α, β, and γ, respectively. The dynein intermediate chains (78 and 69) are shown and eight light chains are indicated by small circles. The ODA-DC is not shown in this figure but is attached to the ICs (Ibanez-Tallon 2003).

Figure 1.6: *Chlamydomonas* **flagellar Inner Dynein Arm (IDA)**
The IDA is more variable than the ODA; there are at least 8 HCs organized into seven isoforms; either double headed light orange) composed of 2 HC (1α and 1β), 3 IC (97, 138 and 140) and three LCs, or single headed (darker orange) of which the exact isoforms has not yet been established. The IDAs interact with the radial spokes via the dynein regulatory complex (DRC), which is not shown in this figure but is located in close proximity on the outer microtubule to the base of the IDA (Ibanez-Tallon 2003).

The composition of the inner dynein arm (IDA) (Figure 1.6) is more variable and is less well understood, as they are more heterogeneous and more complex than the outer dynein arms and are difficult to image in electron microscopy (EM). Three inner arm structures (I1, I2 and I3) are thought to exist and these are distinguished on the basis of their HC composition and their location relative to radial spokes (Piperno et al., 1990). IDAs that are located at the base of the second radial spoke in the axoneme are in close association with the dynein regulatory complex (DRC) that is thought to mediate signals between the radial spokes and the dynein arms (Gardner et al., 1994).

Multiple isoforms of human dyneins are produced by different genes rather than alternative splicing of a number of particular genes, and the functional specificity of dyneins is achieved by the assembly of different isoforms of dynein subunits (Milisav, 1998). Using database searching and analysis it was predicted 12 genes are closely related to *Chlamydomonas* inner dynein arm subunits (Pazour et al., 2006).

Immunofluorescent studies using probes to stain tubulin and heavy chain dyneins DNAH5, DNAI1 and DNAH9 has shown that dynein arm composition differs between cilia and sperm flagella and that it is not uniform along the length of the axoneme (Fliegauf et al., 2005).

1.1.4.2.2 *MAPs: The radial spoke complex*

The radial spokes (RS) are ubiquitous components of 9+2 cilia and flagella. They are T-shaped structures with a "head" and a "stalk" component (Figure 1.7). They are attached by the stalk end to the A-microtubule of the peripheral doublets and reversibly bind via the head to projections of the central pair (Goodenough and Heuser, 1985; Warner and Satir, 1974). Radial spokes are present in pairs or triplet groups located at periodic 96 nm intervals, in exact measure with the inner dynein arms, along the length of the ciliary axoneme (Warner and Satir, 1974). They play an essential role in the control of dynein arm activity by transducing signals from the central pair to the dynein arms via the dynein regulatory complex (DRC), although the exact mechanism is not yet understood (Wemmer and Marshall, 2004). Flagellar movement is inhibited when radial spokes are missing or there are defects in the

central pair (Smith and Lefebvre, 1997). However, the central pair-radial spoke interactions are thought to propagate the cilia bend and direction and the radial spoke-inner dynein arm interactions are thought to play a role in the velocity of the bend (Brokaw et al., 1982). In addition, it has been shown that ciliary motility is partially regulated by phosphorylation and that the central pair-radial spoke complex controls the phosphorylation of the dynein arms, thereby modulating the flagellar waveform (Walczak and Nelson, 1994).

The *Chlamydomonas reinhardtii* flagella radial spoke complex was shown to contain at least 23 proteins, 18 of which make up the stalk and 5 make up the head (Yang et al., 2006) (Figure 1.7). In *Chlamydomonas* it has been observed that lack of one of the radial spoke head subunits results in a complete breakdown of the head complex and therefore absence of the five spoke head proteins within flagella (Yang et al., 2006). A proportion of the protein subunits have Ca^{2+}, AKAP and nucleotide-binding domains that are thought to function in signal transduction. Among the 18 *Chlamydomonas* spoke proteins characterised so far, 12 have apparent homologues in humans, suggesting that this component has been conserved at the molecular level throughout evolution. The importance of the radial spoke in ciliary and flagellar motility is highlighted by the paralysed or abnormal flagellar motility phenotypes observed in *Chlamydomonas reinhardtii* mutants lacking the entire radial spoke complex or just the spoke head.

Figure 1.7: Model for the radial spoke structure.
A) Diagrammatic representation of a cilia cross section demonstrating the 9+2 arrangement of microtubules and the location of MAPs.
B) Diagrammatic representation of the *Chlamydomonas* radial spoke complex, illustrating probable locations of the 23 radial spoke proteins (RSP1-23), based on studies of mutants and analysis of the spoke complex. * indicates newly identified RSPs. Peripheral microtubule doublet is on the left, one of the central pair on the right. RSPs 18 and 19 are not placed due to insufficient evidence indicating their location within the radial spoke. Radial spoke structure from (Yang et al., 2006).

1.1.5 Ciliary beat pattern and mucociliary clearance

Respiratory tract cilia beat in a coordinated manner in order to generate a series of waves which are essential for continuous mucociliary clearance that is an important host defence mechanism against inhaled pathogens (Chilvers and O'Callaghan, 2000). It is generally thought that respiratory cilia have a rigid forward effective stroke which penetrates the mucus layer and thrusts it forward, followed by a backward recovery stroke during which the cilium bends under the mucus layer and repositions itself for the next beat (Cosson, 1996). Analysis of rabbit trachea epithelial cilia revealed that during the recovery stroke, the cilium sweeps sideways and backwards in a clockwise direction that propagates adjacent cilia to beat, causing a ciliary metachronal wave (Sanderson and Sleigh, 1981). Analysis of human respiratory tract cilia movement in two planes using slowed down digital high speed video imaging showed that these cilia beat forwards and backwards (Figure 1.8) with a maximal deviation of less than 5°, with an average beat frequency of 11 Hz (Chilvers and O'Callaghan, 2000). Nodal cilia movement is discussed in Section 1.1.7.

Figure 1.8: Diagrammatic representation of the side view of human respiratory cilia beat pattern.
Adapted from Chilvers and O'Callaghan (2000).

1.1.6 Intraflagellar Transport (IFT)

IFT is the process by which cilia and flagella are assembled and maintained using bi-directional transport of molecular motors and their cargo proteins. IFT was first observed as motility that was unrelated to dynein-based flagella motility in the axoneme of *Chlamydomonas* (Kozminski et al., 1993). Kozminski *et al.* observed particles of varying sizes moving in the anterograde and retrograde direction (Kozminski et al., 1993). It is known that the anterograde movement of IFT particles is powered by the kinesin-II molecular motor (heterotrimeric kinesin, KIF3) (Cole, 1999), whereas the retrograde transport is powered by the cytoplasmic dynein 1b/2 motor (Porter et al., 1999) (Figure 1.9). IFT could also provide a mechanism whereby the cell senses the presence or absence of its cilia and therefore controls transcription of required ciliary genes (Sloboda, 2002). In addition, it is thought to act in signal transduction by transporting extracellular signals to the cell body (Sloboda, 2002). It is at the flagella tip complex (FTC) that cargo loading and unloading, and motor protein regulation occur (Sloboda, 2005) (Pedersen et al., 2006).

Figure 1.9: Diagrammatic representation of Intraflagellar Transport (IFT).
Taken from (Badano et al., 2006) IFT particles (blue) are transported along the outer microtubule doublets (OM) under the ciliary membrane towards the ciliary tip (anterograde) by kinesin and back to the basal body (retrograde) by cytoplasmic dynein.

1.1.7 The role of motile cilia in left-right axis determination

The leftward flow of extra-embryonic fluid has been shown to be critical for the determination of laterality in fish, rabbit and mouse embryos (Hirokawa et al., 2006). This leftward flow is thought to be produced by motile cilia in the mouse embryonic node, present transiently at ~7.5 days postcoitum (Nonaka et al., 1998). This subset of cilia at the node lack the central pair and have a 9+0 structure (Bellomo et al., 1996) (Table 1.1 and Figure 1.10B). They have a circular beat pattern and all rotate in a clockwise direction that creates the leftward flow of fluid that is thought to induce breaking of body symmetry and determine laterality (Okada et al., 2005; Nonaka et al., 1998). KIF3B-deficient mice lack nodal cilia and exhibit randomised left-right asymmetry, unlike wild type controls that had nodal cilia and normal body situs, thus providing evidence that this nodal flow is required for determination of correct body asymmetry during development (Nonaka et al., 1998). Exactly how nodal flow is interpreted to create left-right asymmetry is not completely understood, but it has been postulated that the leftward flow produces an asymmetric distribution of Ca^{2+} and exogenously induced proteins and the creation of morphogen gradients (Okada et al., 2005; Nonaka et al., 1998; Takeda et al., 1999; McGrath and Brueckner, 2003). The nodal vesicular parcel (NVP) model predicts that vesicles filled with morphogens (such as Sonic hedgehog or retinoic acid) are transported leftwards by the nodal flow and smashed open by force to release their contents (Tanaka et al., 2005). It has been hypothesised that a subpopulation of immotile cilia that are abundant near the edges of the node detect this Ca^{2+}/morphogen gradient either by chemical or mechanical stimulation and induce downstream signalling cascades that break bilaterality (Tabin and Vogan, 2003) (Figure 1.10A). Similar laterality breaking mechanisms have also been proposed for zebrafish, involving the Kupffer's vesicle (Essner et al., 2005).

Situs inversus (complete organ reversal) occurs in approximately 50 % of patients with the ciliary disease, Primary Ciliary Dyskinesia (PCD) (Afzelius, 1976) and a further 6 % of cases demonstrate heterotaxy (*situs ambiguous*) giving rise to more complex phenotypes including cardiac abnormalities (Kennedy et al., 2007). Laterality defects in PCD are proposed from model organisms that reflects a loss of nodal ciliary function during embryogenesis.

Figure 1.10: Cilia generate and sense nodal flow for left-right body axis determination during embryogenesis.
A) Adapted from (Yost, 2003). Motile cilia on the cells in the centre of the node (white rectangles) generate fluid flow that is detected either by mechanostimulation or chemostimulation of sensory cilia at the periphery of the node (blue ovals). The fluid flow is proposed to cause a change in Ca^{2+} gradient or to transport nodal vesicular parcels (NVPs) (white circles) that are smashed open by force to release morphogens at the edge of the node and this is detected by sensory cilia. B) Taken from (Hirokawa et al., 2006). The ultrastructure of mouse nodal cilia as viewed by electron microscopy (EM) and schematic representation. Nodal cilia lack the central pair of microtubules so have a 9+0 arrangement.

PART II: CILIA AND DISEASE

1.2 *Ciliopathies*

Cilia are present on almost all polarised cell types of the human body. Dysfunction of cilia can affect many different organ systems, resulting in cilia-related disorders known collectively as the ciliopathies (Figure 1.11). Although all ciliopathies arise from defective cilia, the range of symptoms can vary significantly (Table 1.2). Ciliopathies can affect either single organs or can exist as multi-organ disorders which are phenotypically variable and overlap with other disease manifestations (Badano et al., 2006), including exencephaly (Katsanis, 2006), retinitis pigmentosa (RP) (Krawczynski and Witt, 2004), nephronophthisis (Saunier et al., 2005) and polycystic kidneys (Pazour, 2004), as well as disorders associated with developmental defects such as Bardet-Biedl syndrome (BBS), oral-facial digital syndrome type I (OFD1) and Meckel-Gruber Syndrome (MKS types 1-3).

Non-motile, sensory cilia and are thought to detect environmental signals by various cilia-specific receptors, ion channels and signalling molecules that are present in the ciliary membrane. Signalling through receptor-dependent pathways such as sonic hedgehog (SHH), platelet-derived growth factor receptor (PDGFR) and non-canonical Wnt pathways are mediated through cilia and defects in these pathways can cause developmental defects (Fliegauf et al., 2007). Loss of the Hedgehog (Hh) signalling pathway can cause several birth defects, including polydactyly, craniofacial defects and skeletal malformations (McMahon et al., 2003), which resemble the developmental defects observed in IFT-mutant mice (Zhang et al., 2003). Therefore, cilia-related signalling pathways mediated by IFT are thought to contribute to embryonic development.

Bardet-Biedl syndrome (BBS) is a disease associated with ciliary dysfunction. BBS affects development and is commonly associated with features including cystic kidneys, polydactyly, mental retardation and obesity. Twelve BBS proteins have been identified to date, *BBS1-12* (Table 1.2) and all of those tested have been shown to localise to ciliated cells and tissues. Mice with defective proteins homologous to human BBS1, BBS4 and BBS6 exhibit phenotypes resembling those observed in the

mutants of the non-canonical Wnt pathway. These phenotypes include open eyelids, neural tube defects and disrupted cochlear stereociliary bundles (Torban et al., 2004b; Torban et al., 2004a). Studies in zebrafish and mouse revealed that the protein *LTAP* (also known as *VANGL2*) is involved in Wnt signalling and colocalises with BBS proteins to the basal body and cilia axoneme. Thus providing further evidence that cilia are involved in Wnt signalling and therefore ciliary dysfunction can result in developmental defects.

A similar example is Alstrom syndrome (ALMS), in which patients present with a number of phenotypes similar to those observed in BBS patients, including RP, obesity and diabetes, but also develop deafness and do not have polydactyly. ALMS is caused by mutations in *ALMS1*, a gene which was identified by breakpoint analysis of an ALMS patient with a familial balanced reciprocal chromosome translocation (Hearn et al., 2002), and the protein was shown to localise to both the centrosome and basal bodies in a pattern very similar to BBS proteins (Hearn et al., 2005).

Oral-Facial-Digital syndrome type 1 (OFD1) is an X-linked disorder characterised by malformations of the face, oral cavity and digits with polycystic kidneys and in some cases defects of the central nervous system. OFD1 was shown to be caused by mutations in *OFD1* (Ferrante et al., 2001). The OFD1 protein localises to both the centrosomes and basal bodies, suggesting that this disease may also be related to cilia dysfunction.

Further evidence that ciliary dysfunction contributes to neural tube defects is provided by the fact that mutations in *MKS1* and *MKS3*, which encode the ciliary components MKS1 and meckelin (Dawe et al., 2007; Smith et al., 2006), are associated with a disorder associated with neural tube defects (encephaloceles), Meckel-Gruber syndrome (MKS types 1-3). Recent studies in *Xenopus laevis* embryos have shown that vertebrate Wnt pathway proteins Inturned (Int) and Fuzzy (Fuz) are essential for ciliogenesis (Park et al., 2006), and MKS proteins are implicated in this pathway (Smith et al. 2006).

Animal models have supported the link between cilia dysfunction and renal cyst formation. In addition to the *Tg737*^{orpk} mouse model of autosomal recessive

polycystic kidney disease (ARPKD), in the congenital polycystic kidney (*cpk*) mouse, cystin, the protein product of *cpk*, is present in the cilia of renal epithelial cells (Hou et al., 2002). In autosomal dominant PKD, (ADPKD), mutations have been found in *PKD1* and *PKD2* (Mochizuki et al., 1996). The products of these two genes are polycystin 1 and 2, respectively, which interact with each other and are thought to be part of a Ca^{2+} channel localised in the primary cilium of renal epithelial cells. It has been suggested that both proteins function as mechanosensors of extracellular fluid flow that function by regulating Ca^{2+} flux and that renal primary cilia might act as environmental sensors to regulate cell growth and differentiation. Therefore, dysfunction of these cilia results in abnormal cell proliferation and consequent cyst formation (Nauli et al., 2003). Nephronophthisis (NPHP) is an autosomal recessive cystic kidney disease, for which five genes have been cloned (*NPHP1-5*). Analysis of their protein products has provided a strong link between cilia function and the pathogenesis of the disease (Hildebrandt and Zhou, 2007).

Vertebrate photoreceptors are polarised sensory neurons composed of an inner and an outer segment connected by a highly specialised 9 + 0 connecting cilium. All components necessary for the assembly, maintenance and continuous turnover of the outer segment are transported via the connecting cilium by IFT. In mice, targeted mutations that affect IFT cause retinal degeneration (Marszalek et al., 2000). A link between human retinal degeneration and dysfunction of the connecting cilium has been suggested due to the fact mutations in *RPGR* (Retinitis Pigmentosa GTPase regulator) account for 20 % of cases of retinitis pigmentosa (RP) and the protein localises to the connecting cilium (Hong et al., 2003). In addition, retinal degeneration is a feature of BBS (Katsanis et al., 2000) and as previously mentioned, all BBS proteins are thought to have cilia-related function.

These syndromes all affect sensory cilia function. This study will focus on a disease caused by the dysfunction of motile cilia, including those lining the respiratory tract, that are essential for mucus transport and airway clearance, Primary Ciliary Dyskinesia (PCD).

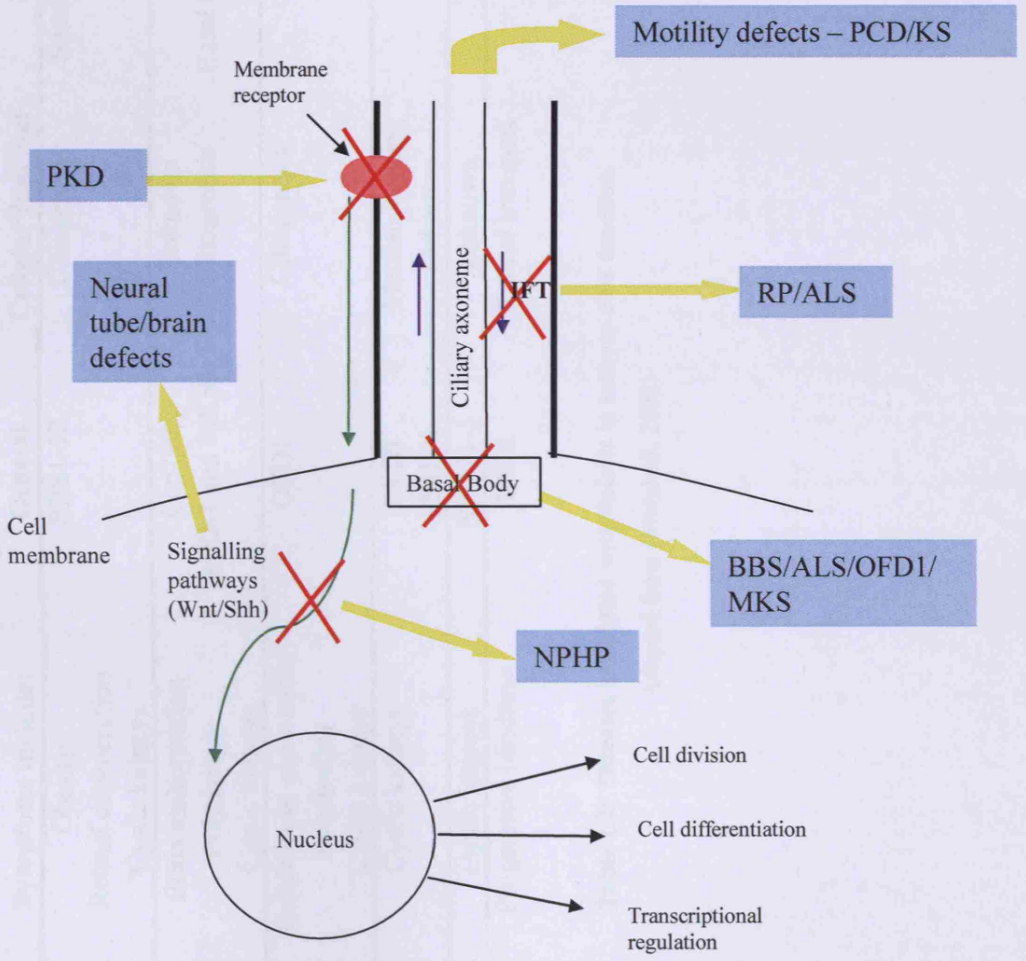


Figure 1.11: Phenotypic outcomes of ciliary defects.
Defects of cilia motility lead to Primary Ciliary Dyskinesia (PCD) and Kartagener Syndrome (KS). Defects in specific receptors in the ciliary membrane cause polycystic kidney syndrome (PKD). Defects in IFT have been linked to retinitis pigmentosa (RP). Perturbation of basal body proteins results in structural and functional or IFT defects caused by dysfunctional signalling cascades, which then can cause a variety of syndromes, nephronophthisis (NPHP) or pleiotropic disorders such as Bardet-Biedl syndrome (BBS), Alstrom syndrome (ALS), oral-facial digital type 1 syndrome (OFD1) or Mekel-Gruber syndrome (MKS).

Table 1.2: Diseases associated with defects in sensory cilia function.

Adapted from (Marshall, 2008)

1.2.1 Primary Ciliary Dyskinesia (PCD)

Primary Ciliary Dyskinesia (PCD) is a genetic condition caused by inherited defects affecting motile cilia function. PCD patients have a range of ciliary defects in a number of tissues and severity of the disease ranges from mild symptoms to severe disease requiring lung transplantation (Carretero Gracia et al., 2000). The pathogenesis of the disease phenotype reflects the distribution of defective respiratory epithelial cilia that results in ineffective mucociliary clearance of the upper airway causing sinusitis, and the lower airways causing chronic inflammatory damage (bronchiectasis). Defective sperm flagella and dysmotility of cilia lining the fallopian tubes causes male or female subfertility, respectively. In addition, dysmotility of the motile cilia in the ependymal cell layer surrounding the brain ventricles, results in a reduction in cerebrospinal fluid flow and occasionally causes the development of hydrocephalus associated with PCD (Picco et al., 1993; Jabourian et al., 1986; Kosaki et al., 2004). Defects of nodal ciliary motility can also lead to left-right symmetry defects (usually *situs inversus*) and left-right axis determination is randomised in patients diagnosed with PCD. In addition, PCD has been associated with other abnormalities including heterotaxy and congenital heart disease (Bush, 1998) (Kennedy et al., 2007) and an association between motile cilia dysfunction and retinitis pigmentosa (RP) and a mental retardation syndrome (OFD1) has also been made (Moore et al., 2005; Budny et al., 2006).

1.2.1.1 History of PCD

Initially, patients with a triad of symptoms, bronchiectasis, sinusitis and *situs inversus* were diagnosed as having Kartagener syndrome (KS) (OMIM 244400). Further research showed that this applied to only a subset of individuals with ciliary defects since not all had *situs inversus* and thus the term immotile cilia syndrome was used (Afzelius, 1976). However, cilia were not necessarily immotile but were sometimes dysmotile or had reduced motility and a residual amount of ciliary function was observed in some patients. Therefore the term Primary Ciliary Dyskinesia (PCD) was introduced in 1981 as an all-encompassing term to describe those people with abnormalities in the structure and function of cilia, resulting in sinusitis and bronchiectasis, either with or without laterality defects (Sleigh, 1981).

1.2.1.2 Clinical presentation of PCD

Clinical symptoms are very variable among PCD patients, ranging from severe neonatal distress and ensuing disease severity, to a delay in diagnosis sometimes until adulthood. A lack of coordinated ciliary movement in the respiratory tract results in reduced mucociliary clearance, causing in sinusitis, rhinitis and/or otitis media, as well as recurrent bacterial respiratory infections leading to chronic progressive inflammatory lung damage and the development of bronchiectasis (Bush et al., 2007). The first presentation of the disease may be unexplained tachypnea in newborn babies, neonatal pneumonia, respiratory distress syndrome, chronic rhinitis, mirror-image reversal of body organs (*situs inversus*) and/or rarely, hydrocephalus. In some cases, the ciliary defect may become more obvious in cases of male subfertility due to dysmotile or immotile spermatozoa flagella (Munro et al., 1994). Female subfertility due to dysmotile cilia lining the oviduct is less frequently observed (Afzelius and Eliasson, 1983).

Approximately 50 % of PCD patients present with *situs inversus* and this reflects the dysfunction of motile cilia in the embryonic node (Section 1.1.7). However, it has recently been shown that at least 6.3 % of patients with PCD have heterotaxy (or *situs ambiguous*, where situs cannot be determined) associated with more severe defects including cardiovascular abnormalities (Kennedy et al., 2007). In fact, the prevalence of congenital heart disease is 200-fold higher in PCD patients than the general population, so it has been suggested that people with chronic coronary heart disease (CHD) should be screened for PCD (Kennedy et al., 2007).

Although the beat frequency of nasal cilia was shown not be associated with brain abnormalities (Roth et al., 1988), there are a number of patients who have been identified with hydrocephalus and dysmotile cilia (Picco et al., 1993). In fact, mice with mutations in the PCD gene *DNAH5* (Section 1.2.1.6), have severe hydrocephalus due to reduced ependymal flow, which is associated with the closure of the cerebral aqueduct during brain development (Ibanez-Tallon et al., 2004).

In a study where 55 cases of PCD were diagnosed, the symptomatic history of patients was documented to highlight the predominant clinical features of the disease (Table 1.3).

Table 1.3: Clinical symptoms of PCD.
Modified from (Coren et al., 2002). 55 cases of PCD were diagnosed and important symptoms documented. Most individuals had multiple symptoms, typically a combination of both upper and lower airway disease.

1.2.1.3 Diagnosis of PCD

Diagnosis is based on the presence of typical clinical phenotypes (Table 1.3) and specific ultrastructural defects of cilia as identified by electron microscopy (EM). Diagnosing PCD is challenging and requires a clinical phenotype that is compatible with clinical tests: ciliary ultrastructural analysis, immunofluorescent staining, ciliary beat assessment, saccharin test and nasal oxide measurements.

Due to limited knowledge about the disease, PCD is often not considered in the diagnosis of patients and as such, diagnosis of PCD is seen to cluster around areas where specialist centres exist. The National Specialist Commissioning Advisory Group (NSCAG) has funded three national diagnostic centres for PCD. The mean age of diagnosis was found to be 4.4 years old in a paediatric PCD clinic in 55 cases, despite the presence of neonatal respiratory distress, *situs inversus* and rhinitis (Coren et al., 2002). However, despite up to three quarters of PCD patients displaying significant symptoms shortly after birth (Bush et al., 1998) (de Iongh and Rutland, 1995) some patients are not diagnosed until later in life.

1.2.1.3.1 *Ultrastructure analysis by EM*

Electron microscopy should ideally be the standard method for diagnosing PCD, in order to identify ultrastructural defects in patients cilia (Bush et al., 1998). However, this is a relatively expensive test and so a broader diagnosis is considered before this test is pursued. In addition, it should be noted that cilia ultrastructure and beat frequency are often disrupted by infection and smoking so cilia ideally need to be cultured in specialist centres in order to get rid of the secondary ciliary defects before EM is performed. Different ultrastructural defects responsible for PCD result in predictable dyskinetic or static beat patterns. Recognition of these can help in the diagnostic evaluation of patients suspected of having PCD (Chilvers et al., 2003). The first ultrastructural defects documented by EM showed absent dynein arms of immotile respiratory cilia and sperm flagella (Afzelius, 1976). Since then a number of other phenotypic variations have been noted, including changes in organisation of microtubules within the axoneme and missing radial spokes (Schneeberger et al., 1980). The most common ultrastructural defect remain that of the outer dynein arms that can now be related to genetic mutations in at least two dynein genes (Noone et al., 2004). For a proportion of cases, no defect can be found at the EM level.

1.2.1.3.2 *Ciliary beat testing*

Ideally, ciliary motility is also assessed to diagnose PCD (Chilvers et al., 2003). The time taken for a patient to taste saccharin after a drop has been placed on the inferior nasal turbinate is a basic measure of ciliary beat frequency (Palmblad et al., 1984). The slower the movement, the more likely it is that the patient has dysfunctional cilia. This method works on the basis that cilia move the saccharin to the tongue. The recent advent of high resolution digital high speed video imaging has allowed the precise beat pattern of cilia to be analysed and utilised in diagnostic procedures (Schipor et al., 2006).

1.2.1.3.3 *Nitric Oxide (NO) testing*

In PCD, nasally exhaled nitric oxide (NO) is unexplainably low (Bodini et al., 2008; Noone et al., 2004). NO is thought to be involved in both the regulation of ciliary motility and host defence. A study by Karadag *et al.* suggested that measurement of nasal NO may be a useful screening test for primary ciliary dyskinesia (Karadag et

al., 1999) and this has been introduced into most PCD clinics. It has been shown that the unaffected parents of PCD patients expected to be heterozygous for the mutation associated with PCD, had lower than expected NO levels in the range between that of normal, healthy subjects and those with PCD (Noone et al., 2004).

It has been noted that caffeine intake may cause a significant decrease in exhaled NO (Bruce et al., 2002). It has been shown that inhalation of L-arginine increases NO in ciliated epithelial cells and leads to a faster CBF (Loukides et al., 1998). However, this is considered a justified tool in the diagnosis of PCD (Wodehouse et al., 2003) and is a routine test used in clinics, although it is not possible to test children younger than about 5 years of age (Piacentini et al., 2008).

1.2.1.3.4 *Immunofluorescence testing*

Detection and intracellular localization of heavy chain dynein, DNAH5, has been observed in respiratory tract epithelial cells by immunofluorescent microscopy along with another heavy chain dynein DNAH9 (Fliegauf et al., 2005). PCD patients with mutations in *DNAH5* either had no DNAH5 staining or mis-localised staining present at the basal bodies, implying that the protein is unable to enter the axoneme. This approach is not in routine clinical use, but could be used in diagnosis of PCD if a panel of antibodies directed towards the products of known PCD genes is developed and may one day present a quicker diagnostic method than EM. Patients with outer dynein arm defects, but no *DNAH5* mutations can also show DNAH5 mis-localisation so a small antibody panel could be developed to detect most axonemal defects.

1.2.1.3.5 *Mutational analysis of candidate genes*

Mutational analysis has demonstrated that an estimated 28 % of PCD families with outer dynein arm defects carry mutations in the *DNAH5* gene and that mutations in the *DNAI1* gene account for approximately 13 % of PCD cases (Hornef et al., 2006a). Therefore mutational analysis of these genes in PCD patients with corresponding ultrastructural defects is offered, although only in the USA at present, and only for detecting common mutations (Section 1.3.2.2). Although the genetics of PCD are slowly being revealed, the multiple phenotypes and the more than 200 proteins

present in the ciliary axoneme mean that it is likely to be some time before such genetic tests are widely used in diagnostic examination. This is partly for economic reasons, especially as dyneins are very large genes.

1.2.1.4 Ultrastructural defects in PCD patients

Electron micrographs (EMs) of cilia or sperm flagella reveal a number of possible ultrastructural abnormalities that cause PCD (Table 1.4) (Afzelius, 1985). These include absence or deficiency of outer and/or inner dynein arms, shortened dynein arms, absent inner dynein arms and nexin links (causes disorganisation of the axoneme), absent radial spokes (the central pair are off centre), absence of one or both of the central pair (although this can be caused by some medications), defect of orientation of the central pair, absence of microtubules in the axoneme, absent axoneme, or abnormal ciliary function despite a normal axonemal ultrastructure (Palmblad et al., 1984). Mutations in the gene encoding a heavy chain dynein, *DNAH11*, were shown to be associated with a normal axonemal structure in a German PCD family with dysmotile respiratory cilia (Schwabe et al., 2008) and a patient with immotile respiratory cilia (Bartoloni et al., 2002). The most common defects (96% of PCD patients reported) involve either or both the outer and inner arms (Afzelius, 1981) (Horvath et al., 2005). In addition, abnormal orientation or length of cilia has been reported (Afzelius et al., 1985). Secondary abnormalities, such as microtubule defects and compound cilia can arise as a result of infection, smoking or pollution and occur in up to 10% of a normal person's cilia, which can confuse the diagnosis of PCD (McAuley and Anand, 1998) (Meeks and Bush, 2000).

Different ultrastructural defects observed by EM have been shown to be responsible for different cilia beat patterns in PCD, observed by digital high speed video imaging (Chilvers et al., 2003). Those with either outer or inner and outer dynein arm defects had a high proportion of immotile cilia, or cilia with only a slight flickering beat pattern. PCD patients with an abnormal ciliary beat pattern of a stiff forward stroke with reduced amplitude, were shown to be have an inner dynein arm defect or a radial spoke defect associated with an inner arm defect. Two patients with radial spoke defects had a circular beat pattern and a higher than normal beat frequency. A ciliary transposition defect was associated with an oval gyrating pattern. Recognition of

such beat patterns might aid in the diagnostic evaluation of patients suspected of having PCD.

Table 1.4: Overview of the ultrastructural abnormalities of respiratory epithelial cilia from PCD patients.
(Adapted from Meeks, 1998 and Afzelius 2004)

1.2.1.5 Treatment and prognosis of PCD

No treatments exist as yet to correct cilia dysfunction, however, management of the disease is important for the clinical course and includes enhancement of mucociliary clearance by chest percussion, aerobic exercise and bronchodilators in some patients (Bush et al., 1998). Antibiotic treatment given promptly upon respiratory tract infection or sinusitis can prevent lung damage. A tympanotomy can help children with ear infection and sinus drainage can relieve symptoms in the short term (Palmblad et al., 1984). Patients with severe bronchiectasis and end-stage lung disease are considered for lung transplantation if respiratory failure looks likely (Palmblad et al., 1984).

Most patients with PCD have a normal life span, although the rate of disease progression can vary and morbidity, rather than mortality, can be considerable if PCD is incorrectly managed (Bush et al., 1998). Patients diagnosed later on in life and who had therefore received less treatment than those diagnosed during childhood, found that PCD had a greater impact on their health in the second half of their life, resulting in significant morbidity and restriction of a normal lifestyle (McManus et al., 2003). Therefore, early diagnosis and treatment would improve symptoms and impact of the condition on patients. More widespread knowledge would lead to an earlier diagnosis and therefore improve the prognosis of PCD patients.

1.2.1.6 Model organisms of PCD

A naturally occurring phenotype similar to PCD occurs in pigs and dogs, except that hydrocephalus is a common feature (De et al., 2004) (Roperto et al., 1993) (Roperto et al., 1994). Naturally occurring or transgenic mouse mutants with features of PCD exist, which include *iv/iv, inv/inv* and *hpy/hop* mutants (Brueckner et al., 1989) (Handel and Kennedy, 1984) (Yokoyama et al., 1993). In *iv/iv* mice, the *lrd* (for left-right dynein) gene was found to be mutated, which is the human homologue of which is *DNAH11* encoding outer dynein heavy chain 11. *iv/iv* mice have a 50 % incidence of laterality defects and immotile nodal cilia, however no respiratory cilia abnormalities have been identified and as previously mentioned *DNAH11* is not yet associated with a specific ciliary defect in PCD patients. The *inv* murine model has

100 % laterality defects, although it also has no ciliary abnormalities and further work is required before the role of this gene is fully understood. The *hpy* and *hop* mouse models have respiratory cilia and sperm flagella abnormalities, thus the males are infertile, however no laterality defects are present. The gene responsible for these phenotypes has not yet been cloned but is thought to map to chromosome 6 (were the cytogenentic band is syntenic to human chromosome 12) (Bryan, 1983).

A mouse *Dnah5* knock out model has been created by gene targeting and this shows a phenotype very similar to PCD patients, in that they develop respiratory problems because of low mucociliary clearance and half have laterality defects (Ibanez-Tallon et al., 2002). In addition the *Dnah5* mice have hydrocephalus unlike the *DNAH5* patients, likely due to a difference in brain structure between species.

The WIC-Hyd rat model is shown to have an X-linked recessive inheritance and these mice display respiratory and ependymal ciliary defects, as well as *situs inversus* (Torikata et al., 1991). More recently, the *Foxj1* knockout mice with a mutation in the forkhead box transcription factor, have been shown to lack 9+2 cilia, whereas 9+0 cilia are present in the embryonic node (Zhang et al., 2004). This is due to lack of transcription of essential basal body components involved in anchoring the basal body to the cytoskeleton and results in a failure of axonemal formation (Gomperts et al., 2004). As a result, mice show random determination of the left-right axis (Tamakoshi et al., 2006) and ciliary genes are down-regulated.

Mice homozygous for the nm1054 mutation in the gene encoding the novel primary ciliary dyskinesia protein 1 (Pcdp1) develop phenotypes associated with PCD (Lee et al., 2008). Homozygous mutants present fatal perinatal hyrocephalus. Heterozygotes have an accumulation of mucus in the sinus cavity and male infertility, due to slow respiratory cilia beat frequency and absence of mature sperm flagella, respectively. Immunohistochemistry showed that the Pcdp1 protein localises to sperm flagella, respiratory epithelial cell cilia and brain ependymal cells in both mice and humans (Lee et al., 2008). This model is therefore an important tool for the study of the molecular genetics and pathogenesis of PCD.

1.2.1.6.1 *Zebrafish as a model organism*

Zebrafish are used as vertebrate disease models primarily because of their experimental and genetic amenability. Zebrafish embryos develop with a transparent chorion and development can therefore be optically observed. In addition, they produce many offspring (up to 200 eggs per female fish per week) and development is rapid. The cost of keeping fish is significantly lower than keeping mice. Therefore, many experiments can be performed on a single clutch and live manipulations and observations can be carried out from fertilisation.

Zebrafish are useful in a number of genetic manipulation experiments, including gain- and loss-of-function techniques (Holley et al., 2002). RNA injection can be performed at particular times and locations in the embryo to assess gene function. Alternatively, modified antisense oligonucleotides, or morpholinos, can be injected into zebrafish embryos to knock down the function of genes by steric blocking of RNA splicing or translation and thereby model a specific human gene defect. A number of zebrafish mutants have been produced by genetic manipulation that are good models of human disease (Lieschke and Currie, 2007), including ones with cilia defects displaying phenotypes such as hydrocephalus and cystic kidneys (Kramer-Zucker et al., 2005). Targeted knockdown of zebrafish heavy chain dynein (*dhc9*) results in the formation of kidney cysts due to reduced pronephric fluid flow (Sullivan-Brown et al., 2008). In addition, zebrafish have a fluid-filled organ called the Kupffer's vesicle (KV), which is analogous to the human embryonic node and is involved in left-right axis patterning (Essner et al., 2002). The similarities between the phenotypes observed in zebrafish caused by defects in motile cilia and the clinical presentation of PCD in humans mean that zebrafish can serve as a useful vertebrate model system for understanding cilia motility and PCD in humans.

PART III: GENETICS OF PCD

1.3 *Inheritance of PCD*

The incidence of PCD is estimated at 1 in 20,000 live births (Meeks and Bush, 2000), although difficulties in diagnosis and the overlap in clinical phenotype with other chronic airway diseases suggests that this is an underestimate (Noone et al., 2004). PCD is an autosomal recessive (AR) disorder (Afzelius, 1998) (Bush et al., 1998), although rare cases of X-linked or autosomal dominant (AD) PCD have been reported (Krawczynski and Witt, 2004) (Moore et al., 2005) (Narayan et al., 1994). As for most AR disorders, the incidence of PCD is increased in isolated, inbred populations. It is genetically heterogeneous condition, in agreement with cilia complexity and the fact that a defect of any of the >250 ciliary components could theoretically result in dysmotile cilia.

1.3.1 Identification of candidate genes for PCD

1.3.1.1 Genetic linkage analysis for positional cloning

Disease loci with a Mendelian pattern of inheritance can be mapped using polymorphic genetic markers such as microsatellites (di-, tri- or tetra- nucleotide repeats) and single nucleotide polymorphisms (SNPs) within affected families. A disease locus is positioned in the genome by being closely linked to genetic markers that have been genotyped at regular intervals across the whole genome (Botstein et al., 1980) and are not shared between affected and non-affected individuals within a single family pedigree.

1.3.1.2 Homozygosity mapping

Homozygosity mapping is an efficient strategy for the identification of positional candidate human genes that cause recessively inherited traits by the identification of genotypes of informative DNA polymorphisms, such as microsatellites and SNPs, in affected offspring of consanguineous families (Lander and Botstein, 1987). With the

complete sequencing of the human genome (Venter et al., 2001) (Lander et al., 2001) a large number of microsatellites and SNPs have been identified for use in this technique.

The homozygosity mapping method works to identify how the genetic information and therefore disease loci have been inherited (Shen et al., 2005) (Puffenberger et al., 2004), based on the fact that affected offspring of a consanguineous marriage will have inherited a region on the chromosome spanning the disease locus that is homozygous by descent (HBD) (the same ancestral disease chromosome from each parent) (Teebi and El-Shanti, 2006). A region HBD that is shared amongst affected offspring and not unaffected offspring may be considered as containing the disease locus (Farrall, 1993). One affected individual from a first cousin marriage has been shown to contain the equivalent genetic information to a non-consanguineous family with three affected siblings (Ott, 1999). The power of a consanguineous family resource therefore allows the number of individuals required for genotyping to be reduced and makes it possible to map rare recessive diseases for which it is difficult to collect adequate numbers of nuclear families with affected offspring and where heterogeneity exists. However, problems with this technique may arise if more than one region of excess homozygosity is identified (since 1/16 of the genome is HBD in a first cousin marriage (Farrall, 1993) and genotyping must be performed on the parents in order to ensure marker informativity (Miano et al., 2000).

The identification of the PCD gene, *DNAH5*, resulted from a genome-wide linkage analysis using homozygosity mapping, and mutational analysis of the dynein heavy chain gene within the candidate interval (Omran et al., 2000).

1.3.1.3 Statistical analysis of linkage analysis

Genetic pedigree data generated by homozygosity mapping is analysed using linkage analysis programmes such as GENEHUNTER (Kong and Cox, 1997) or MERLIN (Abecasis et al., 2002), which calculate LOD scores, using an approximate multipoint calculation that ignores the unlikely possibility of a large number of recombinants. The LOD score is the logarithm of the odds that the loci are linked rather than segregate randomly with a recombination fraction of 0.5, i.e. it is used to calculate the

probability that the pedigree arises randomly or by genetic linkage (Morton, 1955). To be significant, it is generally accepted that a LOD score ≥3 must be achieved.

1.3.1.4 Structural components of cilia as candidate genes for PCD

Many candidate genes exist for PCD, several of which have been unsuccessfully screened and several in which mutations have been identified (Section 1.3.2.2). Model systems for PCD have provided us with a number of candidate PCD genes, as described in Section 1.2.1.6., however, the human genes for these candidates have not all been identified. The structural components of the ciliary axoneme and basal body and the genes that control their expression and assembly are good candidates for PCD. For example, *DNAI1*, the gene for an outer dynein arm intermediate chain, is a homologue of the Chlamydomonas *IC78* and on a candidate gene basis was found to be mutated in a number of patients with outer dynein arm defects. Also, *DNAH11* found to be mutated in a putative case of PCD, was screened on a candidate gene basis since it is the human homologue of the mouse *lrd* gene in the *iv/iv* model. Since the most common ciliary ultrastructural defects seen in PCD patients affect the inner and/or outer dynein arms, each of which is made up of a number of dynein subunits, dynein genes are considered as good candidates genes for PCD. The identification of homologues of *Chlamydomonas* radial spoke proteins in humans indicates that this component is conserved at the molecular level and therefore provides new candidate genes for PCD in patients with radial spoke defects.

Defects in radial spokes are known to cause PCD (Sturgess et al., 1979) (Antonelli et al., 1982) although no gene(s) have yet been identified as disease-causing in patients lacking all or part of the radial spokes. Among the 18 *Chlamydomonas* radial spoke proteins characterised so far, 12 have apparent homologues in humans, suggesting that this component has been conserved at the molecular level throughout evolution (Yang et al., 2006). The importance of the radial spoke in ciliary and flagellar motility is highlighted by the paralysed or abnormal flagellar motility phenotypes observed in *Chlamydomonas reinhardtii* mutants lacking the entire radial spoke complex or just the spoke head. A more complete understanding of the assembly, components and function of the radial spokes in the ciliary axoneme would assist in the identification of such genes.

1.3.1.5 Candidate genes for PCD in the genome

1.3.1.5.1 *Genomic analysis of cilia*

Based on the hypothesis that the ancestral eukaryote was a ciliated organism and that cilia, and therefore cilia-related genes, have been lost throughout evolution in some species, Avidor-Reiss *et al* used a comparative genomics approach to predict the genes required to build cilia. It was assumed that during evolution, genetic information of biological processes and organelles that no longer exist is not preserved. By comparing the genomes of ciliated and non-ciliated organisms a number of proteins involved in the biogenesis and maintenance of the structure and function of cilia were identified. Genomes from organisms that use compartmentalised ciliogenesis (where the basal body docks to the cell membrane and the axoneme and ciliary membrane buds out from this) were compared to the genome of organisms that use cytosolic ciliogenesis (where the axoneme is first assembled inside the cytosol and is then extruded). Using *Drosophila* as a model system, a novel family of proteins involved in compartmentalised ciliogenesis (OSEGs) were reported (vidor-Reiss et al., 2004).

In a similar approach, the proteome of the non-ciliated *Arabidopsis* was subtracted from the proteome of the ciliated organisms *Chlamydomonas* and *Homo sapiens* (Li et al., 2004). These comparisons yielded 688 proteins in the flagella apparatus and basal body (FABB) proteome. These were validated as ciliary components against known *Chlamydomonas* genes, transcriptional regulation of flagella genes and by use of RNA interference techniques (Li et al., 2004).

1.3.1.5.2 *Transcriptomic analysis of cilia*

Using oligonucleotide microarrays, Stolc et al analysed genes induced during regeneration of flagella in *Chlamydomonas* (Stolc et al., 2005). When the *Chlamydomonas* flagella is removed and allowed to regenerate, transcription of flagella structural genes is up-regulated 3 to 12 fold. Genes that are up-regulated upon deflagellation are proposed as likely to be structural components of the

flagellum and also genes whose products are important for regulation of flagella assembly. This approach showed that only 85 out of a total of 220 genes induced had previously been implicated with cilia or flagella and the rest had not previously been connected with flagella function. Any genes that were up-regulated after deflagellation are knocked out using RNA interference and the subsequent phenotype examined (e.g. slow swimming, paralysis, or absence of flagella). Five of the six non-up-regulated genes were shown by RNA interference to cause flagella defects.

1.3.1.5.3 *Proteomic analysis of cilia*

In a proteomic study of human bronchial epithelial cilia, 2D/1D-PAGE and sequencing of proteins by liquid chromatography/tandem mass spectrometry was used along with analysis using Lys-C digestion and subsequent sequencing. Over 200 potential axonemal proteins that matched ESTs were identified, including known PCD genes *DNAH5* and *DNAI1*, as well as the homologue of the zebrafish polycystic kidney disease gene *qilin* (Ostrowski et al., 2002).

Studies of the *Chlamydomonas* flagella predicted that the axoneme contains 360 true and 292 potential flagella proteins (Pazour et al., 2005), this indicated that cilia contain more than the approximated 250 proteins than originally thought and had higher complexity.

An RNAi based interrogation of the *Trypanosome* flagella proteome (TbFP) resulted in defective flagella in this species. In addition, comparison of the TbFP with that of other eukaryotes revealed a number of *Trypanosome* specific proteins. Some genes in the TbFP are homologues of known ciliary disease genes and 34 of the genes were mapped to known ciliary-disease loci, thus are considered potential disease causing candidates (Broadhead et al., 2006). Therefore the proteomic analysis of cilia can be used in the identification of candidate genes for PCD.

1.3.2 The history of PCD gene and loci identification

1.3.2.1 PCD loci

The extensive genetic heterogeneity underlying PCD corresponds to the clinical heterogeneity and the complex structure and regulation of cilia function (Leigh, 2003). Several different loci have been reported to be linked PCD and respiratory defects caused by motile cilia dysfunction have been found in conjunction with other defects such as polycystic kidney disease (PKD) and Senior Loken syndrome (Otto et al., 2005). It has been suggested that it is possible in some families that PCD might be inherited as a digenic disorder, whereby two mutations in two different genes cause the PCD phenotype, however this is yet to be proven (Blouin et al., 2000).

An early genome wide linkage scan on 31 multiplex PCD families with a variety of ultrastructural defects carried out by Blouin et al. suggested no major single gene locus, but several putative PCD loci were suggested on chromosomes 3p, 4p, 5p, 8q, 16p, 17q and 19q (Blouin et al., 2000), thus providing further evidence of the predicted heterogeneity of the disease.

In 1992, affected individuals from 2 unrelated families with Immotile Cilia Syndrome (ICS) with an ultrastructural defect of random microtubule arrangement in the cilia axoneme, were found to share a PCD susceptibility locus within the HLA complex on chromosome 6 (Bianchi et al., 1992).

A total genome scan was performed at University College London (UCL) in five PCD families from Saudi Arabia, with similar ultrastructural defects of absent outer dynein arms, and a putative PCD locus was defined on chromosome 19q13.3 (Meeks et al., 2000). This region is gene rich and contains a large cluster of zinc finger genes. Mutations in the zinc finger transcription factor, *ZIC3*, were been previously reported to cause X-linked laterality defects (Gebbia et al., 1997). However, although a number of chromosome 19q13.3 candidate genes sequenced in the linked families in this laboratory, no mutations have yet been identified.

Another UCL genome-wide linkage scan was carried out on two PCD families from two different, isolated populations: the Faroe Islands and the Israeli Druze population in Israel (Jeganathan et al., 2004). The use of families from isolated and therefore inbred populations, with similar ultrastructural PCD phenotypes reduces the risk of genetic heterogeneity (Miano et al., 2000) (Teebi and El-Shanti, 2006). The Faroe Island families showed ultrastructural defects of absent outer dynein arms (one patient out of eleven had *situs inversus*). The Druze families had a partial loss of inner dynein arm (six out of seven patients displayed *situs inversus*). Three out of four Druze families were linked to a locus on 15q13.1-15.1 and all seven out of eight Faroe Island families were linked to chromosomal locus 16p12.1-13.1, thus demonstrating locus heterogeneity again, despite being very isolated populations. No obvious genes associated with axonemal function were found in the regions identified and this study emphasises the difficulties in mapping PCD genes caused by the underlying genetic heterogeneity.

In a genome-wide search for a common locus in 52 Kartagener Syndrome (KS) families and 18 PCD families (no *situs inversus*), significant linkage to a 3.5cM region on chromosome 15q24-25, with a maximum LOD score of 4.34, was found in 60 % of the KS (but not the PCD) families (Geremek et al., 2006). This locus has since been refined to 1.8 Mb and the 18 genes within the region and 3 nearby were sequenced for mutations in 7 KS probands (Geremek et al., 2008). Sixty SNPs were identified, none of which were thought to be disease causing.

1.3.2.2 PCD genes

1.3.2.2.1 *DNAI1*

The first PCD gene identified was *DNAI1* on chromosome 9p13-21, the gene encoding outer dynein arm intermediate chain 1 (Table 1.5). Two loss of function *DNAI1* mutations were found in a patient lacking outer dynein arms (Pennarun et al., 1999). This gene was excluded in 5 families with similar phenotypes, providing evidence of locus heterogeneity. Guichard et *al* studied 3 independent patients and 2 of their siblings who had PCD and *situs solitus* (Guichard et al., 2001). One of these patients and his brother were heterozygous for a splice defect caused by a T-insertion at position 3 of the intron 1 donor site (IVS1+2_3insT), which prevents splicing and results in a truncated protein. This is the most common PCD mutation in *DNAI1* and has occurred in 4 out of 8 patients, who also have laterality defects. The patients also had another mutation in which a transition mutation (G→A) occurred at position 1543 in exon 16 (Guichard et al., 2001). This resulted in a conserved glycine (seen in *DNAI1* homologues *Chlamydomonas IC78* and sea urchin *IC1*), that possibly acts as an N-myristoylation site, being changed for a serine. Other mutations reported in this gene include insertion of 4 nucleotides in exon 5 which causes a frameshift, a 12bp deletion in exon 17 which truncates the protein and removes a conserved phenylalanine at position 556. In addition a G→C transversion at position 1874 which causes a missense mutation and a G→A transition at 1875, which causes a nonsense mutation, are both present at a conserved tryptophan residue in a WD repeat motif and is thought to result in incorrect folding of the protein (Guichard et al., 2001).

In a separate study, *DNAI1* was found not to be linked in 5 out of 7 unrelated PCD families with an outer dynein arm defect, based on the results of intragenic SNPs and DNA sequencing (Zariwala et al., 2001). In this study, two novel mutations at codon 568 were detected in *DNAI1* – a nonsense mutation leading to a truncated protein and a missense mutation at the same conserved residue (Zariwala et al., 2001). Hornef et al demonstrated that *DNAI1* mutations account for approximately 13 % of all cases of PCD with reported outer dynein arm defects (Hornef et al., 2006a). Zariwala et al identified *DNAI1* mutations, including 12 novel mutations, in 9 % (16 out of 179

families) PCD patients. Most of the *DNAI1* mutations identified occurred as a common founder mutation (IVS1+2_3insT), representing a mutation "hot spot" (Hornef et al., 2006b).

1.3.2.2.2 *DNAH5*

A second PCD gene has been identified; *DNAH5,* which encodes a heavy dynein chain of the outer arm. A total genome scan was performed on an Arabic consanguineous PCD family with absent outer dynein arms and a gene locus was mapped to chromosome 5p15-p14 (Omran et al., 2000). *DNAH5*, the human homologue of the *Chlamydomonas* dynein heavy chain γ, was located within this chromosomal region and was therefore screened as a candidate for PCD. It is expressed in the lung, kidney, and to a lesser extent in the brain, heart, testis and mouse embryonic node between days 7.0 and 8.25 (which explains the *situs inversus* phenotype) (Olbrich et al., 2002). Sequence analysis identified mutations causing defects in the *DNAH5* product (Olbrich et al., 2002). Four homozygous and six heterozygous mutations were identified, most of which caused premature termination resulting in loss of the microtubule binding domain. *DNAH5* mutations are the most common PCD causing mutation discovered to date, where it has been estimated that 28 % of PCD patients yield *DNAH5* mutations (Hornef et al., 2006a).

Mutations in the *Chlamydomonas DNAH5* orthologue created a slow swimming phenotype. A genotype-phenotype relationship has been reported for *DNAH5*, since electron microscopy showed that a translation termination mutation causes a complete absence of the outer dynein arm, whereas splice site mutations cause a partial loss of the outer arm (Kispert et al., 2003). Using specific immunofluorescent antibodies, Fliegauf et al. showed that DNAH5 (and DNAH9) has a regionally restricted distribution along the axoneme. In 3 PCD patients with *DNAH5* mutations and outer dynein arms defects, DNAH5 was completely absent from respiratory cilia and instead accumulated at the microtubule organising centres (Fliegauf et al., 2005). In contrast to respiratory cilia, sperm flagella from a patient with *DNAH5* mutations had normal outer dynein arm distribution, thus demonstrating tissue specific dynein arm composition with differences between cilia and flagella (Fliegauf et al., 2005).

1.3.2.2.3 *DNAH11*

Another heavy dynein chain gene, *DNAH11* was identified as a PCD-causing gene based on a candidate gene rather than a linkage approach. A missense mutation in the homologous mouse gene (*lrd/Dnah11*) was reported in the *iv/iv* mouse, which has laterality defects. *Dnah11* is expressed in the embryonic node at day 7.5, and mutations in human *DNAH11* have been identified in patients with PCD and/or immotile respiratory cilia (Bartoloni et al., 2002; Schwabe et al., 2008).

A mutation screen was conducted in this 82 exon gene in a patient displaying uniparental disomy of chromosome 7 (UPD7). The patient had *situs inversus*, immotile respiratory cilia and cystic fibrosis, the latter due to the presence of a homozygous *CFTR* gene F508del mutation on chromosome 7. A homozygous nonsense mutation (R2852X) was found in exon 52 of *DNAH11*, 10 amino acids before the fourth P-loop causing loss of 3 of the six AAA domains (section 1.1.4.2.1) (Bartoloni et al., 2002). The N-terminal tail is still present, so the protein is predicted to correctly incorporate into the dynein arm but not to bind to the adjacent microtubule, therefore preventing cilia bending. This prediction is consistent with the electron microscopic data of the patient in which the axonemes and dynein arms appear normal. It should be noted that there is some controversy over whether this patient really suffers from PCD or whether the symptoms are a result of cystic fibrosis and the *situs inversus* is just a coincidence.

Sequence analysis of *DNAH11* in an additional 6 selected PCD sibships that shared alleles in this gene revealed an R3004Q substitution mutation in a conserved residue in one person, although *Chlamydomonas* α and γ heavy chains have a Q at this position. However, none of the control families analysed had the Q3004 variant so it was presumed possible that the R3004Q mutation is pathogenic (Bartoloni et al., 2002).

Six affected individuals from a German PCD family with immotile cilia but an apparent normal cilia ultrastructure as observed by EM, were found to carry compound heterozygous mutations c.12384C>G and c.13552_13608del (Schwabe et al., 2008). Both mutations are predicted to result in a truncated DNAH11 protein.

Therefore, these *DNAH11* mutations are associated with a PCD with a normal ultrastructure.

1.3.2.2.4 *DNAH7*

The proteome of normal and PCD bronchial cells were compared by Zhang et al. and the dynein heavy chain DNAH7 was found to be absent from the cilia in one patient. Although no mutations were detected in the full-length cDNA clone, the DNAH7 protein was found to be synthesised but not assembled in the cilia of bronchial PCD cells. Instead it accumulates intracellularly, in contrast to normal cells where DNAH7 is clearly present in the cilia (Zhang et al., 2002). Thus, it was hypothesised that PCD in this patient resulted from a defect in the transport of DNAH7 from the cytoplasm to the cilia axoneme, although the actual gene mutation was more likely to reside in another gene since no mutations were found in *DNAH7* in this PCD patient.

1.3.2.2.5 *TXNDC3*

A thioredoxin-nucleoside diphosphate kinase, *TXNDC3*, was recently identified as a gene involved in PCD (Duriez et al., 2007) (Table 1.5). Thioredoxins function as enzymatic proteins involved in protein disulfide reductase reactions, although their exact role in relation to cilia structure or function is not understood. However, expression of *TXNDC3* was shown in testis and respiratory epithelial cells it was also shown that it is the human orthologue of the sea urchin intermediate dynein chain *IC1*, a component of sperm flagella outer arms (Duriez et al., 2007). A patient presenting with a PCD phenotype was found to harbour two defects in *TXNDC3*; a nonsense mutation predicting a truncated protein and a common intronic sequence variant that reduced the amount of one particular isoforms. This was an unusual SNP-induced modification of the ratio of two isoforms of a protein that results in PCD. The individual displayed heterogenous ultrastructural abnormalities of respiratory cilia, with partial lack or shortening of outer dynein arms in 66 % of cilia (Duriez et al., 2007).

1.3.2.2.6 *Genes causing syndromic forms of PCD*

Mutations in the gene for Retinitis pigmentosa GTPase regulator (*RPGR*) have been found in one family with retinitis pigmentosa (RP), hearing loss and bronchosinusitis, and it has been suggested that the *RPGR* gene could be involved in some cases where RP and PCD coincide (Zito et al., 2003). RPGR is present as a structural component in the connecting cilium in the retina, and is also expressed in the lung and testes. 70 % of X-linked retinitis pigmentosa (RP) is caused by a mutation in the *RPGR* gene (located on Xp21.1). Moore et al. investigated a non-consanguineous family in which the mother has RP and her 2 sons had PCD (characterised by partial dynein arm defects) and secondary RP. Sequence analysis of *RPGR* in the 2 boys and their mother revealed a deletion, resulting in a severely truncated protein (Moore et al., 2005).

A large Polish family with X-linked recessive mental retardation (XLMR) syndrome had macrocephaly and ciliary dysfunction with symptoms of respiratory distress. A frameshift mutation was identified in the *OFD1* gene in this family (Budny et al., 2006). This supports the role of *OFD1* in respiratory epithelial ciliary function. Oral-facial-digital type 1 syndrome (OFD1S) is an X-linked dominant disorder affecting development. Further mutations in the *OFD1* gene have been shown to be responsible for OFD1S (Ferrante et al., 2001). *Ofd1* heterozygous knockout female mice were shown to develop kidney cysts were no cilia were present. Heterozygous males were not viable, but embryos had laterality defects, indicating that Ofd1 protein is involved in left-right axis determination (Ferrante et al., 2006).

1.3.2.2.7 *Candidate PCD genes screened with no mutations identified*

Other candidate genes have been sequenced in a number of different PCD families (Table 1.5), but were excluded since no pathogenic mutations were identified, just polymorphic variants. These genes include *HP28*, the human homologue of the *p28 Chlamydomonas* gene (Geremek and Witt, 2004), *hPF20*, the homologue of *Chlamydomonas pf20* (mutants of which have paralysed flagella) (Pennarun et al., 2002) and *TCTE3*, the homologue of murine *Tcte3* (mutants of which have reduced sperm motility) (Neesen et al., 2002), *HFH-4*, a fork-head transcription factor

recently renamed *FOXJ1* (Maiti et al., 2000), *DPCD*, the human homologue of the mouse *Dpcd1*, which is defective in the mouse mutant with hydrocephalus and PCD phenotypes (Zariwala et al., 2004), *DNAI2*, a gene related to *Chlamydomonas* dynein intermediate chain *IC69* (Pennarun et al., 2000) and *DNAH9*, a dynein heavy chain that is mis-localised in the axoneme in patients with *DNAH5* mutations (Bartoloni et al., 2001).

Gene	Mutations detected	% of cases of PCD that mutations are accountable for	Ciliary localisation	Ultrastructural defect
DNAI1	32	13%	ODA	Absent ODA
DNAH5	10	28%	ODA	Absent ODA
DNAH11	3	Not known	ODA	None
DNAH7	None detected	Not known	IDA	Not known
TXNDC3	1 (plus 1 intronic variant)	Not known	Not known in human	Absent or shortened ODA
RPGR	1	Not known	Not known	Partial dynein arm defects
OFD1	1	Not known	Not known	Not known
HP28	None detected	N/A	IDA	Not known
hPF20	None detected	N/A	CP	Not known
TCTE3	None detected	N/A	ODA	Not known
HFH-4	None detected	N/A	Not known	Not known
DPCD	None detected	N/A	Not known	Not known
DNAI2	None detected	N/A	ODA	Not known
DNAH9	None detected	N/A	Not known	Not known

Table 1.5: PCD candidate gene screened for mutation. ODA = outer dynein arm, IDA = inner dynein arm, CP = central pair.

1.4 *Aims*

This study aimed to make progress in the genetic and molecular characterisation of PCD. To conduct genome-wide screens with the aim of identifying genetic loci by linkage analysis and homozygosity mapping in consanguineous PCD families. To undertake fine mapping of regions where novel loci are identified and eventually to evaluate candidate PCD genes using a comparative bioinformatic approach.

2 CHAPTER TWO

2.1 *MATERIALS*

Chemicals purchased from Sigma unless stated otherwise.

2.1.1 General use laboratory reagents and solutions

0.5M EDTA (ethylenediaminetetraacetic acid)

EDTA.Na_2.$2H_2O$ (pH8.0)	186.12g/L

10X TBE

TRIS base	108g/L
Boric acid	55g/L
EDTA	8.3g/L

TE Buffer

TRIS-HCl (pH8.0)	10mM
EDTA	0.1mM

Agarose

GibcoBRL electrophoresis grade (for 2% gel)	20g/L

APS

Ammonium persulphate (10% w/v)	10g/L

Ethidium Bromide

Ethidium Bromide	10 mg/ml

DNA Ladders

1 Kb ladder	(*Invitrogen*)
1Kb plus ladder	(*Invitrogen*)
HyperLadder V	(*Bioline*)

2.1.2 Loading buffers

Blue Dextran Loading Buffer

De-ionised formamide (99.5% pure)	10 ml
50mM EDTA (pH8.0)	2 ml
30mg/ml Blue Dextran	12 ml

Orange G Buffer

Glycerol:250mM EDTA (pH8.0)	1:1
SDS (sodium dodecyl sulphate)	0.1% w/v
Orange G powder	0.1% w/v

2.1.3 Growth media

Bacterial growth media - Luria-Bertani (LB) Broth

Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L

For solid medium add 1.5 % (w/v) bacto-agar prior to autoclaving

***Chlamydomonas reinhardtii* growth media – TAP media**

X40 TAP concentrate

Tris base	96.8 g/L
Distilled H_20	920 ml
1M (K)PO_4	40 ml (see recipe below)
Glacial acetic acid	~40 ml to pH7.0
Store at 4 °C	

1M (K)PO_4

1 M K_2HPO_4	250 ml
1M KH_2PO_4 (titrate to pH 7.0)	170 ml

4X Beijerinck salts

NH_4Cl	16 g/L
$CaCl_2.2H_20$	2 g/L
$MgSO_4.7H_2O$	4 g/L

Hutner Trace elements

Supplied by Dr. Pinfen Yang, Marquete University. Made according to protocol in Hutner, et al. (1950). Methods in Enzymology, 23:68.

1X TAP media

X40 TAP concentrate	25 ml
Distilled H_2O	949 ml
4X Beijerinck salts	25 ml
Hutner trace elements	1 ml

Autoclave, store at room temperature.

For solid medium, add 2 % bacto-agar prior to autoclaving

2.1.4 DNA extraction buffers

ACD solution

Trisodium citrate	13.2g/L
Citric acid	4.8g/L
Dextrose	14.7g/L

Solution I

Sucrose	0.32M
TRIS-HCl pH7.0	10mM
$MgCl_2$	5mM
Triton X-100	1% w/v
Sodium azide	0.02% w/v

Solution II

TRIS-HCl pH8.0	50mM
EDTA pH8.0	20mM
SDS	2% w/v

Solution III

Saturated NaCl solution

2.2 *Technical laboratory equipment*

Centrifuges	Centrifuge 5415C/5417C	(Eppendorf)
	Sorvall RT 6000 D Du Pont	(Legend)
Camera for analysis of agarose gels	MP4+ Land Camera System	(Polaroid)
Developing Machine	RG11 Fuji X-ray film processor	(Fuji)
Electrophoresis Tank	Hybrid	(GibcoBRL)
Gel Imager	GDS 8000 Documentation System	
UV Transilluminator	Model 750-M	(International Biotechnologies Inc.)
PCR Machines	9700 Thermocycler	(Applied Biosystems)
	Tetrad Thermo-cycler	(MJ Research, Watertown, Mass)

Sequencer/genotyper ABI 373 Prism (ABI)

Incubator for *Chlamydomonas* culture (25 °C)

INNOVA 4340 illuminated shaker incubator (New Brunswick Scientific)

Fume hood K3 class I sterile air cabinet (Bassaire)

Microscopes Leica DM 2500 light microscope (X10 eyepiece, X40 objective lens)

Inverted Nikon Diaphot microscope

High speed video camera

Trouble-shooter 500, Lake Image systems, UK

2.3 *Family Resource*

Two groups of families were studied which were grouped according to their geographical origin and ultrastructural phenotype. Details of patients and clinical diagnosis are shown in Tables 2.1, 2.2 and 2.3.

2.3.1 Pakistani families 120, 130, 141, 143 and 145

Families 120, 130, 143 and 145 were from Mirpur city or Rawalpindi of Northern Pakistan. Family 141 were from the New Frontier area of Pakistan. All five families followed the traditional emigration route to the Bradford area of the UK. All five families were first cousin consanguineous families with a total of nine affected and seven non-affected siblings (Figure 2.1A). DNA was available for all individuals except 145 II:2. All affected individuals had *situs inversus* except for 120 II:3. Electron microscopy (EM) results for these families indicate missing inner and outer dynein arms (Figure 2.1B). High speed imaging analysis of nasal ciliary brush biopsy demonstrated that affected individuals had immotile, or flickering cilia with a significantly reduced beat frequency.

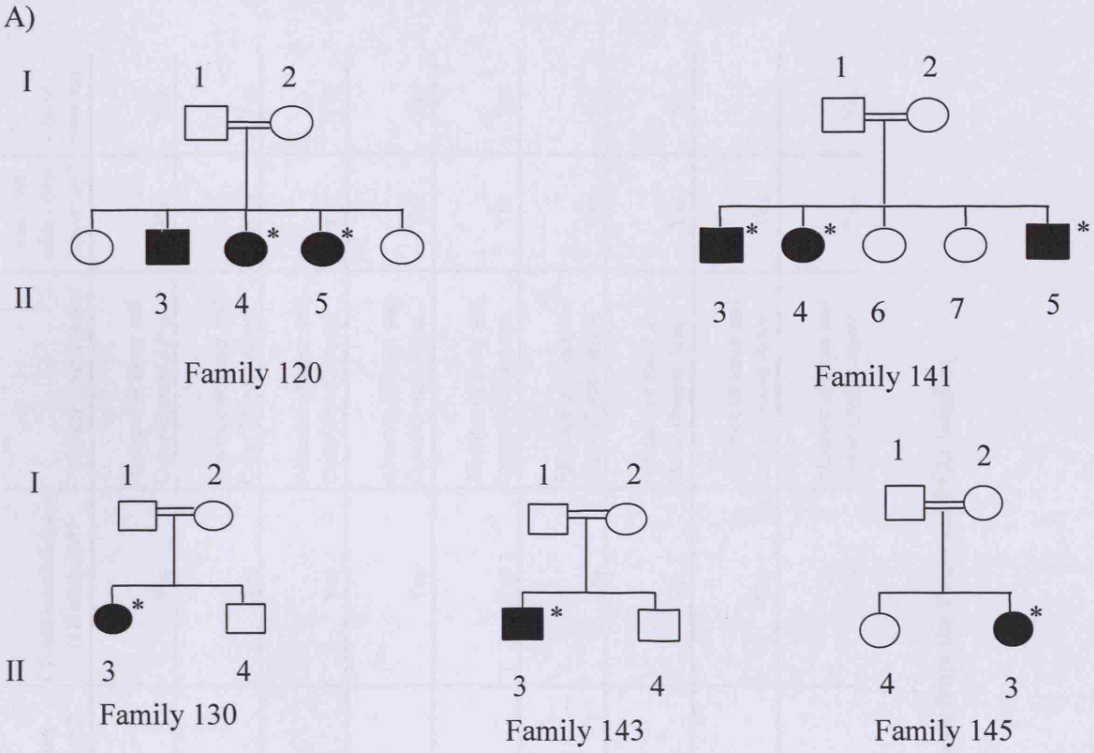


Figure 2.1: Pakistani family pedigree structures and EM analysis. A) Pakistani family group pedigree structures. * indicates affected patient with *situs inversus*. Circles indicate female, squares indicate male. Filled in symbols are indicate individual was affected. Double line indicates consanguinity (first cousin marriage). B) EM from family 120 – I) EM of patient 120 II:3 showing absent inner and outer dynein arms. II) Cartoon of EM with inner and outer dynein arms shown in yellow.

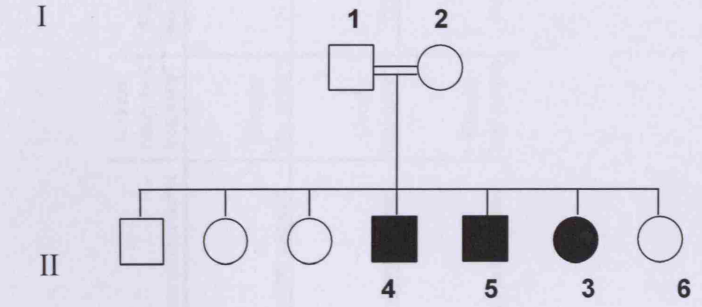
ID	Region	Age of diagnosis	Symptoms	Documented bronchiectasis?	CF/immunodeficiency /TB excluded?	Ultrastructural defect	Reduced ciliary beat frequency?	*Situs inversus*
120 II:3	Mirpur, Pakistan	7 yrs	cough, rhinitis, sinusitis, otitis media, hearing loss, low weight, neonatal respiratory distress	Yes	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
120 II:4	Mirpur, Pakistan	birth	cough, rhinitis, sinusitis, low weight	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
120 II:5	Mirpur, Pakistan	birth	cough, rhinitis, sinusitis	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
130 II:3	Mirpur, Pakistan	4 yrs	cough, rhinitis, sinusitis, otitis media, low weight, neonatal respiratory distress	Yes	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
141 II:3	Rawalpindi, Pakistan	5 yrs	cough, rhinitis, sinusitis, otitis media, low weight, neonatal respiratory distress	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
141 II:4	Rawalpindi, Pakistan	birth	cough, rhinitis, sinusitis, low weight, neonatal respiratory distress	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	No
141 II:5	Rawalpindi, Pakistan	birth	cough, rhinitis, sinusitis, low weight, neonatal respiratory distress	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	No
143 II:3	North West Frontier, Pakistan	3 yrs	cough, rhinitis, sinusitis, otitis media, low weight, neonatal respiratory distress	Yes	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes
145 II:5	Mirpur, Pakistan	3 yrs	cough, sinusitis, low weight, neonatal respiratory distress	N/A	Yes	Absence of Inner and Outer Dynein Arms	Yes	Yes

Table 2.1: Clinical information for affected individuals from the Pakistani PCD families

2.3.2 Arabic family 146

Family 146 is a consanguineous Arabic family (paternal first cousin marriage) from the United Arab Emirates (UAE), originating from an ancient Bedouin tribe called the Bani Tameem which settled in Doha in the early 18th Century. They have three affected and four non-affected offspring (Figure 2.2A). There is no incidence of *situs inversus*. High speed video imaging of nasal epithelia biopsies demonstrate that the cilia have a circular beat pattern and 11.9 - 17 % of the electron micrographs (EM) of cilia cross sections show an unusual absence of the central pair or "central pair agenesis" (Stannard et al., 2004) (figure 2.2B). This is a significant reduction compared to the level of absent central pairs seen in normal, non-affected cilia (0-0.2%). It has been proposed that the central pair appears and disappears at 96 nm periodicities along the length of the axoneme (Prof. C. O'Callaghan, personal communication) (Figure 2.2C). Their cilia beat frequency was normal (Stannard et al., 2004).

A)



C)

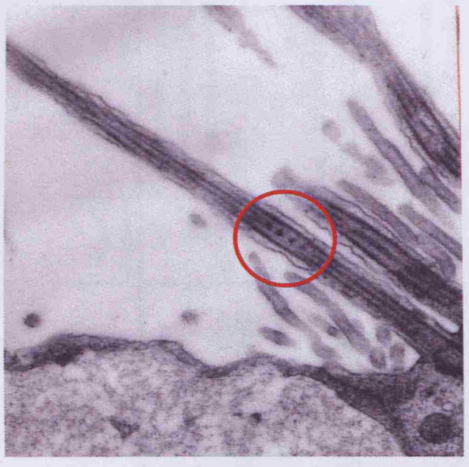


Figure 2.2. Arabic Family 146 pedigree structure and EM analysis. A) UAE pedigree structure, circles indicate female, squares indicate male. Filled in symbols are affected. Double line indicates consanguinity. B) EM from UAE family and controls - i) normal 9 + 2 cilia EM. Ii) central pair agenesis seen in the UAE family, missing central pair (9 + 0 structure). Taken from *Stannard et al, (2004) Am J Respir Crit Care Med, 169,634-637.* C) Longitudinal section respiratory cilia EM from affected patient from family 146 (Unpublished data from Prof. C. O'Callaghan, University of Leicester. Intermittent central pair is highlighted with a red circle.

ID	Region	Age of diagnosis	Symptoms	Documented bronchiectasis?	CF/immunodeficiency /TB excluded?	Ultrastructural defect	Reduced ciliary beat frequency?	*Situs inversus*
146 II:3	United Arab Emirates	5 yrs	Cough, reduced exercise tolerance, rhinorrhea, low weight, short stature, bronchiectasis, collapsed lower left pulmonary lobe	Yes	Yes	Central pair agenesis	Circular beat pattern	No
146 II:4	United Arab Emirates	4 yrs	Cough, reduced exercise tolerance, rhinorrhea, low weight, short stature,	N/A	Yes	Central pair agenesis	Circular beat pattern	No
146 II:5	United Arab Emirates	8 months	Cough, reduced exercise tolerance, rhinorrhea, low weight, short stature,	N/A	Yes	Central pair agenesis	Circular beat pattern	No

Table 2.2: Clinical details for Arabic PCD family 146

2.3.3 Arabic family 152

Family 152 is a complex Arabic family of Bedouin origin although their precise tribal origin was not known. All marriages within this pedigree were first-cousin in origin. There were five affected and twenty-eight non-affected individuals in total (Figure 2.3). Non-affected individual 152 II:6, with four affected siblings, had one affected and eight non-affected offspring. This is an apparently low incidence compared to the generation above. One of the four affected siblings in family 153 had died at the time of clinical ascertainment due to an unrelated condition. The father of individual 152 I:1 had two wives who were his cousins. In addition, the two wives were thought to be related but the exact nature of this relationship was uncertain so was left out in this pedigree. As a result, the exact details were uncertain for the degree of consanguinity between 152 I:1 and I:2. There is no incidence of *situs inversus* in this family. Ultrastructural analysis of respiratory cilia by EM revealed an apparently normal ultrastructural, although it was considered likely that the defect was missed at this level. Using high speed video imaging of nasal brush biopsies a circular ciliary beat pattern with normal CBF was observed for this family. This was consistent with the ultrastructural defect of intermittent central pair seen for family 146.

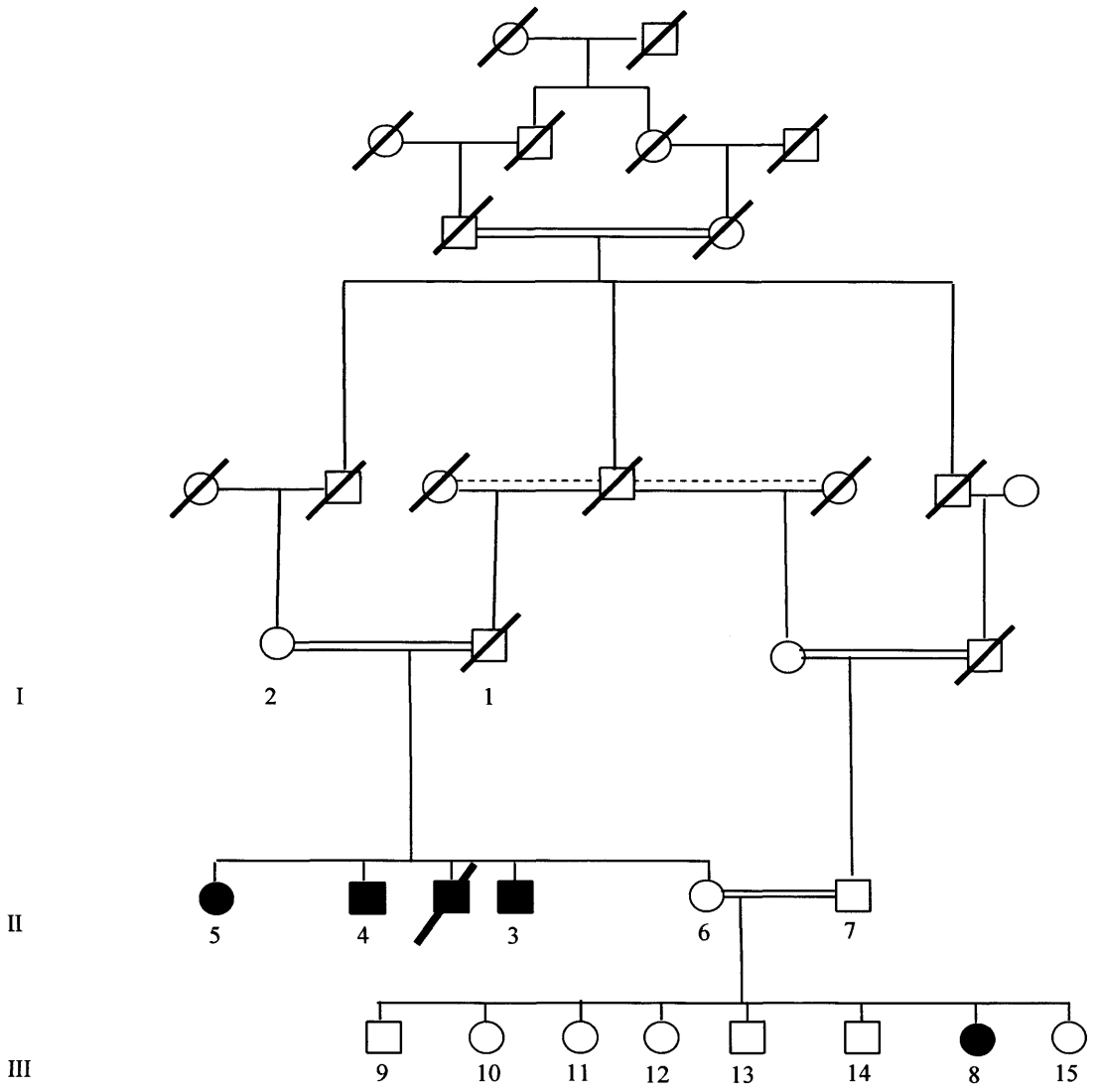


Figure 2.3. Arabic family 152 pedigree structure. Circles indicate female, squares indicate male. Filled in symbols are affected. Double line indicates consanguinity. Crossed out symbols are deceased. Dotted line indicates cousin marriage of uncertain nature.

ID	Region	Symptoms	CF/immunodeficiency/TB excluded?	Ultrastructural defect	Reduced ciliary beat frequency?	*Situs inversus*
152 II:3	Arabic Bedouin	Cough, reduced exercise tolerance, intermittent rhinorrhea, low weight, delayed mucociliary clearance, poor sperm motility.	Yes	None observed	Circular beat pattern	No
152 II:4	Arabic Bedouin	Cough, reduced exercise tolerance, intermittent rhinorrhea, low weight, delayed mucociliary clearance, poor sperm motility.	Yes	None observed	Circular beat pattern	No
152 II:5	Arabic Bedouin	Cough, reduced exercise tolerance, intermittent rhinorrhea, low weight, delayed mucociliary clearance. Had IVF due to infertility.	Yes	None observed	Circular beat pattern	No
152 III:8	Arabic Bedouin	Cough, reduced exercise tolerance, intermittent rhinorrhea, low weight, delayed mucociliary clearance	Yes	None observed	Circular beat pattern	No

Table 2.3: Clinical details for Arabic PCD family 152.

2.4 *METHODS*

2.4.1 Preparation of genomic DNA and RNA

2.4.1.1 Preparation of genomic DNA from human whole blood

See Section 2.1.4 for buffer details.

5-20 ml blood samples were taken by venepuncture and added to 50-200 µl 0.34M EDTA or acid citrate dextrose (ACD) to prevent the samples from coagulating. The blood samples were then transferred to 50 ml conical tubes and coded to protect the anonymity of patients. The samples were centrifuged at 12,000 x g for 15 minutes at 4°C and the plasma was removed with a pipette. Pelleted blood cells were stored at -20°C until DNA extraction was performed.

Distilled H_2O was added to the samples to a total volume of 50 ml and pellets dispersed by gentle mixing. After incubation on ice for 5-10 minutes, samples were re-pelleted by centrifugation (12,000 x g at 4°C). Supernatant was removed and 40-45mls solution I (1X) added to a total volume of 50mls to cause lysis of blood cells.

After 2 minutes incubation on ice, the sample was centrifuged to separate cell debris in the supernatant from DNA and cell contents present in the pellet. The previous steps detailed were repeated until the pellet was light pink in colour (a red pellet indicated clotted material was still present).

The pink pellet was resuspended in 11 mls solution II by vortexing vigorously. Proteinase K was added to a final concentration of 100 ug/ml, to digest proteins present in the cell and allow them to form complexes with solution II, and the solution incubated at 37°C overnight. The sample was then held on ice for 10 minutes then 4 mls of solution III was added and mixed by invertion. The mixture was then held on ice for a further 5 minutes to allow precipitation of protein complexes, which were then pelleted by centrifugation (12,000 x g for 15 minutes). Supernatant was transferred to a sterile 50 ml tube and RNase added to a final

concentration of 20 mg/ml. The solution was then incubated at 37°C for 15 minutes to digest any RNA present.

Genomic DNA was precipitated by addition of 2 volumes of 100 % ethanol at 4°C to the supernatant, with gentle invertion to mix. The DNA was observed to precipitate and was pelleted by centrifugation at room temperature for 30 minutes. The pellet was then rinsed in 70 % ethanol and air-dried. DNA was then resuspended in 200 µl of TRIS-EDTA buffer at 4°C overnight to allow full resuspension of the DNA. Stock DNA was stored at -20°C and working concentrations of 15 ng/µl were made using sterile dH_2O to dilute and stored at 4°C. 5–10 ml of blood typically yielded 30-225 mg DNA.

2.4.1.2 Preparation of genomic DNA from *Chlamydomonas reinhardtii*

Total DNA was extracted using a method adapted from Goldschmidt-Clermont (1990). A 10 ml aliquot of cells grown to a density of 1 x 10^6 cells/ml was harvested at 6000 g for 5 minutes. The pelleted cells were resuspended in 0.6 ml of 0.2 M Tris (pH 8.0 – HCl), 0.5 M NaCl, 0.01 M sodium EDTA and 0.2 % (w/v) SDS and freeze thawed three times using liquid nitrogen and a 37 °C water bath. The sample was then diluted with 0.6 ml of TE saturated phenol:chloroform:IAA (25:24:1 (v/v) respectively) and centrifuged at 10000 g for 2 minutes. The upper aqueous phase was then diluted in 1.4 ml of 95 % (v/v) ethanol and centrifuged at 10000 g for 10 minutes. The pellet was washed with 70 % ethanol and resuspended in 50 µl of DEPC treated sterile water. DNA was stored at -20 °C until required.

2.4.1.3 Preparation of RNA from zebrafish embryos

Total RNA from zebrafish embryos at 24, 48 and 72 hpf was isolated using 10 to 50 dechorionated embryos per sample. Embryos were homogenised by vortexing after addition of 500 µl TRIzol Reagent, (*Invitrogen*) which maintains the integrity of RNA during lysis. After incubation of the homogenised samples for 5 minutes at room temperature to permit the complete dissociation of nucleoprotein complexes, 0.2 ml chloroform per 1 ml TRIzol was added. The samples were mixed vigorously

and centrifuged at 6000 x g for 15 minutes at 4 °C to separate the aqueous and inorganic phases. The aqueous, upper layer was recovered, mixed with 0.5 ml isopropanol, incubated for 10 minutes at room temperature and centrifuged at 6000 x g for 10 minutes at 4 °C. The pellet was washed once with 70 % ethanol, centrifuged again for 2 minutes, air-dried and dissolved in 50 µl dH_2O. Total RNA was stored at -80 °C.

2.4.2 Preparation of cDNA

First-strand cDNA was synthesised using the Amersham Biosciences First-Strand cDNA Synthesis Kit according to the manufacturer's instructions.

Briefly, the RNA was placed in a microcentrifuge tube and RNase-free water added to bring the volume to 20 µl. The RNA was then heated at 65°C for 10 minutes to denature, chilled on ice then mixed gently. 11 µl of bulk first-strand cDNA reaction mix containing Moloney Murine Leukaemia Virus (M-MulV) reverse transcriptase, RNase-free BSA and dNTPs was added to a clean tube along with 1 µl DTT solution, 1 µl primer (pDN6 random hexamers) and the heat denatured RNA. This was mixed gently and incubated at 37°C for one hour and then stored at -20°C for use in PCR amplification (section 2.4.5).

2.4.3 Quantification of DNA

The concentration of each patient DNA prepared from whole blood (section 2.4.1.1.) was determined spectrophotometrically or using PicoGreen fluorescence.

For spectrophotometer quantification a 1 in 50 dilution of each sample was made and the optical density measured at a wavelength of 260 nm. The concentration was then determined using the following equation:

OD_{260} X thickness of cuvette (50 ug/ml) X dilution factor (50) = DNA concentration (ng/µl)

PicoGreen quantification was determined according to the instructions for the Quanti-iT PicoGreen dsDNA Assay Kit (*Invitrogen*). Briefly, diluted DNA samples were assembled at a dilution of 1:10 (100 µl) with 100 µl of PicoGreen solution in a Costar 96-well black flat bottom assay plate. To create a standard curve, 100 µl of control DNA at concentrations of 100, 200, 400, 600, 800, 1600 ng/µl with 100 µl of PicoGreen was also assembled. The PicoGreen fluorescence emission was measured using a Tecan GENios fluoscan linked to Magellan software. DNA concentrations were quantified against the standard curve and a known control concentration used to check accuracy.

2.4.4 Primers for genotyping and sequencing

Primers for known and novel microsatellite markers were identified using information in the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and supplied by Sigma, Invitrogen or MWG Biotech.

In-house primers were designed using the Primer3 programme (Rozen and Skaletsky, 2000), to a uniform design of approximately 20 bp, a Tm of 60°C and a GC content of approximately 50 %. A 3' GC clamp was incorporated where possible and primer sequences were confirmed to be unique using the BLAST programme. To enable efficient annealing, the primer sequences were designed 100% homologous to the target DNA sequence. To decrease the formation of primer dimers, complementarity within a primer or between both primers was avoided.

Primers for sequencing of genomic exons were designed ≥40 bps from intron-exon boundaries in order to sequence all coding regions plus splice sites. Details of primers for in-house microsatellites and gene sequencing are shown in Appendices.

2.4.5 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) utilises *in vitro* polymerisation to amplify specific DNA sequences (Saiki et al., 1985). Unique oligonucleotide primers were designed to amplify specific regions of DNA.

In general, the following reagents were added to a 0.5ml 96-well microtiter plate (Appleton Woods) or 0.5ml tubes for PCR optimisation: (10X PCR buffer = 50 mM TRIS-HCl (pH8.8 at 25°C, 200mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween® 20) All reagents supplied by Invitrogen, except primers which were ordered from MWG.

Reagent	Final concentration
Template DNA	15 ng
Primer (both)	1 mM
dNTP	0.25 mM
10x PCR buffer	1X
Magnesium chloride (MgCl2)	1 - 3.5 mM
Taq DNA polymerase	0.05 units
sterile distilled water	to volume of 15 µl
Total volume	15 µl

Table 2.4: Reagents used in PCR

The plate was sealed to prevent evaporation. PCR was performed in a DNA Engine Tetrad thermocycler (MJ Research, Watertown, Mass.). Cycling conditions of denaturation, annealing and extension varied according to the T_m of the primers and the length of the DNA being amplified. A negative control (no template DNA) was included for each set of PCR reactions. Optimisation of the reaction was necessary for each primer pair. The magnesium ion concentration was titrated (1-3mM), since, in addition to other reaction components binding magnesium ions, *Taq* DNA polymerase also requires free magnesium ions for its activity. Therefore, the amount of free magnesium may vary for different PCRs. In order to facilitate denaturation of *Chlamydomonas reinhardtii* DNA molecules with a high G/C content, dimethyl sulphoxide (DMSO) was used at 2.5-5% of the total reaction volume. Standard cycling

conditions for PCR were an initial denaturation step at 94°C for 210 seconds, amplification for 35 cycles (35 seconds denaturation at 96°C, 30 seconds annealing at primer Tm, and 60 seconds extension at 72°C per 1 kb of DNA). The final extension step was 120 seconds at 72°C.

2.4.5.1 PCR Optimisation

If products were difficult to amplify, the 10X *Invitrogen* PCR buffer was replaced with 10X BioLine PCR buffer with NH_4^+ and amplified using the appropriate PCR cycle.

The *Invitrogen* PCR Optimiser Kit was used where products were difficult to amplify. 5X buffer solutions were used which yielded a range of different [$MgCl_2$] and pH. Buffers A, B, C and D (pH 8.5 and varying 1.5 mM – 3.5 mM Mg^{2+}), buffers F, J and N (2.0mM Mg^{2+} and pH 8.5 – 10.0) were used first. If further optimisation was required all buffers were re-tested with the addition of the universal solvent, DMSO, which makes primer-binding easier, at a final concentration of 10 %, while adjusting water so that the final volume remained at 25µl. The following reaction was used :

Reagent	Final concentration
DNA	15 ng
Primer (both)	0.2 µM each
dNTP	0.25 mM
10x PCR buffer (including MgCl2 for final concentration)	1X
Taq DNA polymerase	0.08 units
sterile distilled water	to volume of 25 µl
Total volume	25 µl

2.5: Reagents used in optimisation of PCR

The plate was sealed to prevent evaporation. PCR was performed in a DNA Engine Tetrad thermocycler (MJ Research, Watertown, Mass.). Cycling conditions of denaturation, annealing and extension were performed as in section 2.4.5.

2.4.5.2 Nested PCR

If the first round cDNA PCR yielded a weak product and two rounds of amplification were required in order to achieve a visible and specific product, a second round PCR was performed after the initial PCR reaction on first strand cDNA. This second round 'nested PCR' was performed using a second set of internal primers, positioned within the initial desired PCR product. A standard PCR reaction was performed for second round PCR, as described in section 2.4.5 substituting 1 µl of 1:20 diluted first round PCR product as the DNA template.

2.4.6 Agarose gel electrophoresis

Agarose gel electrophoresis was employed to determine the specificity and size of the amplified product of PCR or restriction endonucleases digestion. The size of a DNA molecule determines its electrophoretic mobility in different concentrations of agarose. Larger DNA molecules migrate slower than smaller molecules due to a greater frictional drag and less efficient migration through the gel pores. In general DNA molecules greater than 1kb were size fractionated in a 1% agarose gel. Molecules smaller than 1kb were separated in a 1.5-3 % agarose gel.

Agarose gels were prepared by combining the appropriate amount of powdered agarose (Life Technologies Ltd.) with 100ml 0.5xTBE electrophoresis buffer and subsequently dissolved in a microwave oven. 1µg/ml ethidium bromide was added to the solution, which was stirred and cooled to approximately 70°C. The solution was then poured into a plastic gel-former (sealed at both ends with tape) with a casting comb positioned 0.5-1mm above the tray. The gel was solidified at RT for 30 minutes and placed in a horizontal electrophoresis tank containing 0.5xTBE electrophoresis buffer. Approximately 2µl loading buffer was mixed with each DNA sample and loaded into the wells. A 1Kb (*Invitrogen*) or HyperLadder V (*Bioline*) ladder (3µl) was run adjacent to PCR products in order to estimate their size. The gel was electrophoresed at room temperature at 80-120V until the required separation was achieved. Ethidium bromide is a fluorescent dye that intercalates between the base pairs of a DNA molecule. The bound ethidium bromide fluoresces under ultraviolet irradiation thereby enabling visualisation of the DNA molecules. The products were visualised in UV light using a UV

transilluminator and photographed using a UVP digital image capture system and the Grab-IT 2.04.7 software package (UVP).

2.4.7 Genotyping analysis

Microsatellite PCR products were separated according to allele sizes by polyacrylamide gel electrophoresis using an ABI 373 sequencer. Fluorescent labels attached to the forward primer are visualised by a laser. Different fluorescent labels emit different colours (FAM = blue, TET = green, HEX = yellow) and the size standard Tamra (Applied Biosystems) emits red fluorescent light. 1 µl of diluted PCR product, 3.5 µl Blue dextran formamide loading buffer and 0.5 µl Tamra size standard was denaturated at 95°C for 3 minutes and loaded onto a 6 % polyacrylamide (Sequa-gel 6) gel containing 10 % APS. The gel was run at 40 Watts, 1100 Volts and 40 mAmps for 5 hours. Allele sizes were called and genotypes derived using the ABI-PRISM Genotyper 2.5 software package.

2.4.8 Single nucleotide polymorphism (SNP)-based genome search

SNP genotyping was performed using a commercial service at The Turku Centre for Biotechnology, University of Turku and Abo Akademi University, Finland and the *Illumina* SNP-based Linkage IVb Panel, integrated with Illumina's high-performance assay. DNA samples were sent to the Finnish genome centre and 5,800 SNPs were typed across the entire genome and genotypes sent back for analysis.

2.4.9 Linkage analysis

Parametric linkage analysis was carried out using the MERLIN programme (Abecasis et al., 2002), assuming an autosomal recessive inheritance, a disease allele frequency of 0.007 (based on the Hardy-Weinberg equilibrium) and a disease penetrance of 1.0. Five parameter files are used in this programme: DAT file (lists genetic markers), MAP file (locates genetic distance of markers), FREQ file (lists allele frequency of markers), PED file (contains family and genotype information) and MODEL file

(contains disease information). These files were compiled according to SNP allele information supplied by *Illumina.* Where additional microsatellite data was incorporated into the parameter files for linkage analysis, their positions relative to the SNPs were taken from a genetic map constructed from the *Illumina* deCODE genetic map for SNPs and the deCODE genetic map for microsatellite markers (Kong et al., 2002). Estimated locations of in-house markers were determined with reference to the UCSC browser (http://genome.ucsc.edu).

2.4.10 Identification of PCD candidate genes using bioinformatics

On the basis that cilia are evolutionarily well conserved and to rationally prioritise genes that could be involved in ciliary disease for mutational analysis, a comparative bioinformatics approach was used to identify positional candidate genes with a potential role in the structure or function of cilia, in linked regions of interest identified in PCD families. Ciliate and non-ciliate genome and proteome datasets were used as well as transcripts from the *Foxj1* mouse microarray. *Foxj1* is an f-box transcription factor expressed in ciliated cells and was shown to be required for late stage cilia formation by a gene deletion in the $Foxj1^{-/-}$ mouse model. Microarray analysis demonstrated that genes expected to be involved in ciliogenesis are down-regulated in this mutant (pre-publication data made available by Dr. S Brody, Washington University).

Known and predicted genes within the regions of interest were identified using the UCSC genome browser (http://genome.ucsc.edu) and coding region sequences for predicted genes were then downloaded from Entrez (http://www.ncbi.nlm.nih.gov). These were then subjected to a BLAST search using a cut-off E value for significance of homology of $\leq 10^{-5}$. A PCD candidate gene was expected to be conserved in ciliates but not in non-ciliates. The genomic sequence freeze dates used at the time the grid was designed (July 2005). Searches were performed against whole genome sequences from nine ciliated and three non-ciliated organisms, using a local Unix-based grid designed by Dr. Richard Emes (Centre for Applied Entomology and Parasitology, Keele University) and Dr. Rahul Chodhari (UCL). Human mRNA

sequences were translated in 6 frames and screened against the translated genome (6 frames) or proteomes of the model organisms, using a TBLASTX programme.

The 9 ciliate datasets comprised:

Chlamydomonas reinhardtii genome (genome.jgf-psf.org/chlre2/chlre2.home.html)
Pazour's *Chlamydomonas* Flagellar proteome (Pazour et al., 2005)
Chlamydomonas reinhardtii deflagellation dataset (Marshall WF 2005, Proceedings of the National Academy of Science)
Trypanosoma brucei and *Leishmania major* combined genomes (Ivens et al., 2005; Berriman et al., 2005)
Ciona intestinalis genome
Tetrahymena thermophilia genome (www.lifesci.ucsb.edu/~genome/Tetrahymena/)
Gull & McKean's *Trypanosoma* flagellar proteome (Courtesy of Prof. Gull & Dr. McKean, Lancaster University)
Caenorhabditis elegans proteome (sensory cilia only) (www.sanger.ac.uk/Projects/C_elegans/WORMBASE/current/wormpep_download)
Foxj1 mouse ciliogenesis microarray (genes down-regulated in cilia-less mice compared to controls) (pre-publication data made available by Dr. S Brody, Washington University)

The non-ciliated datasets comprised:

Arabidopsis thaliana (www.arabidopsis.org)
Saccharomyces cerevisae (www.yeastgenome.org)
Saccharomyces pombe (www.sanger.ac.uk/Project/S_pombe/protein_download)

Candidate genes were prioritised for mutational analysis according to the results obtained from the BLAST search, in addition to available information on their human tissue expression pattern from Genbank, NCBI, Ensembl and dbSNP, protein domain structure (NCBI Conserved Domain Database v2.13) and known or predicted

function of the gene products according to published experimental information (PubMed) and NCBI.

2.4.11 PCR purification

2.4.11.1 EXOSAP-IT

PCR products for sequencing were purified using EXOSAP-IT (USB), containing Exonuclease I and Shrimp Alkaline Phosphatase. 2 µl EXOSAP-IT was added to every 5 µl PCR product (removed from mineral oil). This was activated at 37°C for 15 minutes and deactivated at 80°C for 15 minutes, then stored at 4°C.

2.4.11.2 QIAquick PCR purification

The manufacturer's protocol for the QIAquick PCR purification Kit (QIAgen) using a microcentrifuge was followed. All centrifugation steps were performed at 13,000 x g.

5 volumes of buffer PB (containing hydrochloride and isopropanol) was added to 1 volume of PCR sample and mixed. The sample was then added to a QIAquick spin column and centrifuged for 1 minute. Flow-through was discarded and 0.75 ml buffer PE added to wash and centrifuged for 1 minute. Flow-through was discarded in centrifugation was repeated for an additional 1 minute, to remove residual ethanol from buffer PE. The QIAquick column was placed in a clean 1.5 ml eppendorf. DNA was eluted by adding 30 µl elution buffer (10 mM TRIS-HCl pH 8.5), the column left to stand for 1 minute and centrifuged. The average eluate volume was 28 µl, which was stored at -20°C.

2.4.11.3 QIAquick Gel extraction

Where multiple bands were observed after agarose gel electrophoresis, indicating non-specific amplification, the QIAquick Gel extraction Kit (QIAgen) protocol using

a microcentrifuge was used, following the manufacturers protocol. All centrifugation steps were performed at 13,000 x g.

Briefly, PCR products of appropriate size were excised from the agarose gel with a clean, sharp scalpel. The gel slice was weighed and placed in a tube. 3 volumes of buffer QG (containing guanidine thiocyanate) was added to 1 volume of gel and incubated at 50°C until the gel had dissolved. 1 gel volume of isopropanol was added to the sample, which was then added to a QIAquick column and centrifuged for 1 minute. Flow-through was discarded and 0.5 ml buffer QG added and centrifuged for 1 minute, to remove all traces of agarose. The column was washed with 0.75 ml buffer PE and centrifuged. The column was placed into a clean 1.5 ml eppendorf and 30 µl elution buffer added, left to stand for 1 minute and centrifuged for 1 minute. The average eluate volume was 28µl, which was stored at -20°C.

2.4.12 Plasmid DNA purification

Plasmid DNA was extracted from overnight cultures of *E. coli* in LB medium using the QIAprep mini-prep kit (QIAgen) according to manufacturer's instructions. 1 µl of a 1:20 dilution of eluted DNA was used in subsequent PCR reactions.

2.4.13 DNA sequencing

Purified PCR products of a size up to 400 bp were subjected to the following PCR reaction, according to the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) protocol:

50 ng per 300 bp PCR product
1 µl Pellet Paint NF Co-Precipitant (*Novagen*)
5 pmol primer (forward or reverse)
4 µl Sequencing reagent premix (contains Thermo Sequenase II DNA polymerase)
Distilled H_2O to 20 µl total reaction volume

The reactions were covered with mineral oil and amplified using the following reaction on a GeneAmp PCR system 9700 thermocycler:

96°C 20 seconds
50°C 15 seconds
60°C 60 seconds
For 25 reaction cycles

The PCR products were purified according to the Amersham Bioscience DYEnamic ET Terminator Cycle Sequencing protocol: 3 M NaAc and 95 % EtOH were added to the entire PCR reaction, the solution centrifuged and DNA pellets washed with 70 % EtOH and air-dried. Pellets were resuspended in blue Dextran formamide loading buffer, denatured at 95°C for 3 mins and loaded onto a 6 % acrylamide denaturing gel which was run for 12 hours at 1500 V, 30 W and with a laser power of 40 mW. The fluorescence emitted by the dideoxy-termination sequences is read by the laser.

For products larger than 400 bp, purified PCR products after EXOSAP-IT treatment were sent for commercial sequencing to the Advanced Biotechnology Centre, Imperial College, University of London where the ABI Big Dye terminator kit version 3 from PE/ABI and an ABI PRISM 7700 Sequence Detection System are used. Alternatively, samples were sent to Alpha Bio-laboratory Express DNA Sequencing Service, California for sequencing, using Applied Biosystems ABI

3730xl/3130xl fluorescence-based capillary electrophoresis genetic analyzers. Electropherograms data was received by email.

Sequence electropherograms were analysed and aligned to a reference sequence for mutational analysis using the Sequence Navigator (ABI) or GeneTool software (BioTools Incorporated).

2.4.14 In situ hybridisation probe design and synthesis

Mouse embryos were analysed at developmental stages E 7.5, E18 and E19 by *in situ* hybridisation performed by Dr. Nick Greene, UCL ICH.

The mouse *C6ORF206/RSP9* antisense probe for *in situ* hybridisation running from base position 317 in exon 2 to base position 837 in exon 5 was amplified by PCR using a 1:20 dilution of purified mouse *C6ORF206/RSP9* clone (RZPD) and primer sequences 5'-TGGTGAGTGGCCGTTTCAT-3' and 5'-CCATGTTCTTCTCTCCTGTGC-3'. The resulting product was cloned into vector pCR4 by TOPO-TA cloning, according to the manufacturer's (*Invitrogen*) instructions. Antisense and sense probes were synthesized by Dr. Nick Greene, UCL ICH using Sp6 and T7 promoter priming sites.

2.4.15 Maintenance of zebrafish lines

Breeding zebrafish (*Danio rerio*) lines were maintained at 28°C on 14h light/10h dark cycle. Fertilised eggs were obtained from natural spawning and grown in incubators at 28.5^{0}C, 22^{0}C (to slow development) or 31^{0}C (to accelerate development) according to the stages required. Embryos were staged according to standard references (Kimmel et al., 1995).

2.4.16 Obtaining non-pigmented embryos

Formation of melanin was blocked by incubating embryos in 0.2 mM 1-phenyl-2-thiourea (PTU) at 24 hours onwards to prevent pigmentation of developing embryos without affecting growth.

2.4.17 Observation of live embryos

Embryos at somitogenesis stages and later stages of development were visualised in fish tank water and manually dechorionated with # 5 watchmaker's forceps. When necessary, for photographing live, embryos were anaesthetised with 0.02 % tricaine (3-amino benzoic acid ethyl ester) and mounted for viewing in 3 % methylcellulose in fish tank water.

2.4.18 Morpholino targeted gene-knockdown in zebrafish

Morpholino targeted gene knockdown was performed in wild-type control zebrafish, provided by the Department of Anatomy and Developmental Biology, UCL. Morpholinos work through an RNase-H independent process that blocks correct splicing of the transcript.

2.4.18.1 Morpholino oligonucleotide design

Morpholinos were obtained from Gene Tools, LLC. Morpholinos were arbitrarily designed to bind to the RNA spanning exon-intron boundaries of exon 2 – intron 2 and exon 3 –intron 3. Sequences were selected based on design parameters according to the manufacturer's recommendations, namely 21-25mer antisense oligonucleotides of ~50 % G/C and A/T content with no predicted internal hairpins. Four consecutive G nucleotides and representation elsewhere in the genome were also avoided. Sequences were as follows: MOex2 5'- GGTGTAAGGCTTTTACCGTGACCTC-3' and MOex3 5'- GCTGTAAGTATACCTCCAAAGCTTC-3'.

2.4.18.2 Morpholino preparation

GeneTools prepared morpholino-modified antisense oligonucleotides. The supplied 300 nmol of lyophilised powder was diluted as a stock solution in 100ul of 1X Danieau buffer (5mM HEPES pH7.6, 58mM NaCl, 0.7mM KCl, 0.4mM $MgSO_4$,

0.6mM$Ca(NO_3)_2$) to a concentration of 25 mg/ml. The stock solution was diluted to working concentrations of 6.25 mg/ml in 1X Danieau solution before injection. For morpholino distribution analysis, Green Fluorescent Protein (GFP) mRNA was co-injected into the embryos and analysed under blue light. At this concentration, MO-injected cells developed normally. Images were obtained using a Leica DM 2500 light microscope at 300 X magnification and a KODAK DCS420 digital camera.

2.4.18.3 Morpholino injections

Embryos were obtained at the 1-2 cell stage and microinjection targeted to the cytoplasm or yolk syncytial layer. Siblings from the same pool of MO-injected embryos served as the internal controls. Needles were pulled from glass capillary tubes using a Clark Electromedical Instruments needle puller and injections were performed using a Picospritzer micro-injector.

Effective doses were determined for each morpholino. For example, at the effective dose of 6.25 ng and lower, embryos developed normally as assayed under normal morphological criteria. Higher doses of the GFP-MO resulted in a larger average reduction of GFP protein, but also caused some detectable detrimental effects on development (data not shown). These higher-dose effects were not pursued further. In all cases shown, the dose used for analysis resulted in embryos of two classes, those displaying a specific phenotype or those that were normal using morphological criteria. A small fraction (typically less than 5 %) of embryos developed abnormally due to mechanical damage following microinjection, these embryos were discarded before analysis.

2.4.19 Analysis of zebrafish olfactory pit cilia motility

Seven to nine control and MO-injected embryos at 72 hours post-fertilisation (hpf) were analysed individually under a glass coverslip, taking care not to damage the embryo or the yolk. Embryos were confirmed as alive before analysis by checking their heart beat. Olfactory pit cilia were viewed under a water immersion objective lens (x50) on an inverted Nikon Diaphot microscope in a humidified (80 %) and temperature (28°C) controlled chamber. High-speed (500fps) video sequences were

captured (Trouble-shooter 500, Lake Image systems, UK) of the zebrafish olfactory pits. The stored sequences were then replayed in slow motion (Midas 2.0 player, Xcitex) and the cilia beat frequency (CBF) was calculated using the following formula.

Frames per second (500)/ Frames elapsed for 5 beat cycles X beat cycles counted (5) = CBF (Hz).

The Immotility Index (II) and Dysmotility Index (DI) were determined by counting the number of static or dysmotile cilia as a percentage of the total number present in the video sequences.

2.4.20 Growth and maintenance of *Chlamydomonas*

Chlamydomonas reinhardtii were maintained on TAP-agar plates (Section 2.1.3.) at 25 - 28 °C and re-streaked onto fresh plates every month. Cultures were streaked and re-streaked from stock plates or newly grown colonies in order to get a well growing, healthy colony before expanding into liquid culture. A loop-full of culture was taken from TAP-agar plates and re-suspended in liquid TAP media (Section 2.1.3.) and grown for 16 to 72 hours, whilst slowly shaking under light at 25 - 28 °C.

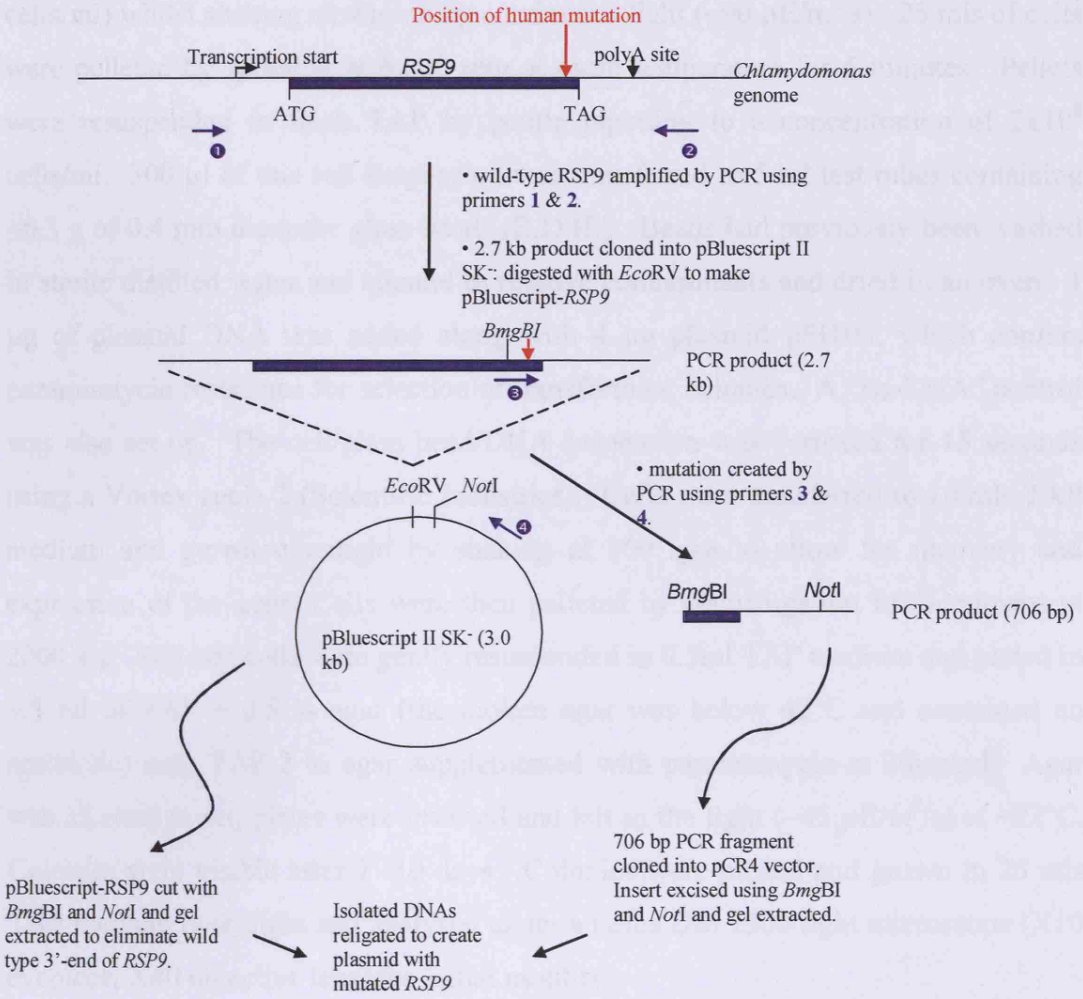
2.4.21 Creation of DNA constructs for wild-type and mutant *Chlamydomonas RSP9* transformation

The wild-type genomic *RSP9* gene was amplified by PCR from wild-type *Chlamydomonas* genomic DNA. The gene sequence amplified included the 5' and 3' UTR and approximately 400 bp of the upstream sequence incorporating the tub box sequence motif (GCTC(G/C)AAGGC) that is known to enhance transcription after deflagellation (Davies and Grossman, 1994). Primer sequences used for genomic PCR were: 5'- AGATTCCACACCTCACGGATAC-3' (Figure 2.4 - P1) and 5'-ACCAGTCAAACTTTCGAACCAG-3' (P2). pFusion *Taq* DNA polymerase (*Invitrogen*) was used in the PCR, which creates blunt-ended products allowing compatibility with *EcoRV* – digested plasmid pBluescript II SK$^-$ (Figure 2.4).

For transformation of *pf17* with mutated *RSP9*, a 3 bp deletion was introduced into this wild-type DNA construct, using a long PCR primer with the relevant three bases missing. This 3 bp change recreates the amino acid deletion identified in the C6ORF206/*RSP9* linked families and the restriction enzyme site *BmgBI* created at the 5' end. This was primer P3:

5'-
CGAGCTGACGTGGGGCAGCCTGTACGTGGGCGACGGCCTGAACAACGAC
C-3'

(Figure 2.4). The reverse primer used was complementary to pBluescript, downstream of a unique *NotI* site, primer P4: 5'–GTATGTTGTGTGGAATTGTGAGCGG-3'. Amplification of pBluescript-RSP9 with P3 and P4 created a 706 bp PCR product. To ensure correct digestion of the ends, the mutagenic PCR was cloned into PCR4 TOPO-TA vector and then excised using *BmgBI* and *NotI* double restriction enzyme digestion and gel purification. Wild-type pBluescript-*RSP9* was digested using the same enzymes and gel extracted to eliminate the wild-type *RSP9* insert. Both isolated DNAs were then ligated together to create a plasmid with mutated *RSP9* (Figure 2.4). Each DNA had a sticky end (*NotI*) and a blunt end (*BmgBI*) that ensured correct orientation of ligation and prevented re-circularisation of vector.



❶	AGATTCCACACCTCACGGATAC P1
❷	ACCAGTCAAACTTTCGAACCAG P2
❸	CGAGCTGACGTGGGGCAGCCTGTACGTG GGCGACGGCCTG|AACAACGACC P3 (|= 3 bp CGC deletion WT, purple = *Bmg*BI)
❹	GTATGTTGTGTGGAATTGTGAGCGG P4

Figure 2.4: Scheme for the synthesis of wild type and mutated constructs used in the transformation of *Chlamydomonas* mutant *pf17*. Left: Table of primers used.

2.4.22 Transformation of *Chlamydomonas reinhardtii*

The protocol used was based on the glass bead method of Karen Kindle (Kindle, 1990). Cells were grown in TAP medium at 25 °C to mid-log phase (1 - 2 x 10^6 cells/ml) whilst shaking slowly under continuous light (~90 μE/m^2/s). 25 mls of cells were pelleted by spinning at 6,000 rpm at room temperature for 5 minutes. Pellets were resuspended in fresh TAP by gentle pipetting to a concentration of 2x10^8 cells/ml. 300 μl of this cell suspension was transferred to 5 ml test tubes containing ~0.3 g of 0.4 mm diameter glass beads (B.D.H.). Beads had previously been washed in sterile distilled water and ethanol to remove contaminants and dried in an oven. 1 μg of plasmid DNA was added along with 4 μg plasmid pSI103, which confers paramomycin resistance for selection of transformant colonies. A "no-DNA" control was also set up. The cell/glass bead/DNA suspension was vortexed for 15 seconds using a Vortex genie-2 (Scientific Industries). Cells were transferred to 10 mls TAP medium and grown overnight by shaking at 100 rpm to allow for recovery and expression of the gene. Cells were then pelleted by centrifugation for 5 minutes at 2000 x g. Pelleted cells were gently resuspended in 0.5ml TAP medium and plated in 3.5 ml of TAP + 0.5 % agar (the molten agar was below 42°C and contained no antibiotic) onto TAP 2 % agar supplemented with paramomycin at 20μg/ml. Agar was allowed to set, plates were inverted and left in the light (~45 μE/m^2/s) at ~22°C. Colonies were visible after 7 -10 days. Colonies were picked and grown in 25 mls TAP medium over night and analysed under a Leica DM 2500 light microscope (X10 eyepiece, X40 objective lens) for initial motility.

2.4.23 Analysis of *Chlamydomonas* transformant motility

100 μl of *Chlamydomonas* suspension grown overnight was placed on a glass cover slip and viewed under a water immersion objective lens (x50) on an inverted Nikon Diaphot microscope in a humidified (80 %) and temperature (30°C) controlled chamber. High-speed (500 fps) video sequences were captured (Trouble-shooter 500, Lake Image systems, UK) of the *Chlamydomonas*. The stored sequences were then replayed in slow motion (Midas 2.0 player, Xcitex) and the flagella beat frequency (FBF) was calculated using the following formula.

Frames per second (500)/ Frames elapsed for 5 beat cycles X beat cycles counted (5) = FBF (Hz).

Chlamydomonas motility was variable in the non-synchronised cultures used, therefore only the FBF was calculated from *Chlamydomonas* when their beat pattern was symmetrical. The Immotility Index (II) was determined by counting the number of static flagella as a percentage of the total number present in the video sequences.

3 CHAPTER THREE

3.1 *LINKAGE ANALYSIS AND MUTATION SCREENING IN FIVE PAKISTANI PCD FAMILIES*

This chapter describes a high density SNP-based genome wide linkage screen, multipoint linkage analysis and a homozygosity mapping approach that were used to identify the causative gene in a group of five PCD families. These families had a high incidence of *situs inversus* (eight out of nine affected individuals) (Figure 3.1 and described more fully in Section 2.3). PCD is a genetically heterogeneous disorder (Blouin et al., 2000) and therefore a large single family or a group of multiple small families with a clear common phenotype or from the same geographical ethnic origin is preferable to increase the chance of genetic homogeneity. The families used in this study were all of the same geographical origin, from the Mirpur region of Northern Pakistan. In addition, they all share a common ultrastructural ciliary phenotype: a defect of absent inner and outer dynein arms. The parents in each family are first cousins and there are a total of nine affected and seventeen unaffected individuals (Figure 3.1). This chapter describes the identification of two genetic loci on chromosome 11q23.3-24.3 and 17q21.31-22 that were statistically significant for linkage in the Pakistani families. To identify the disease-causing gene, postional candidate genes were prioritised for mutational analysis according to their known or predicted function, expression pattern and sequence conservation in ciliates.

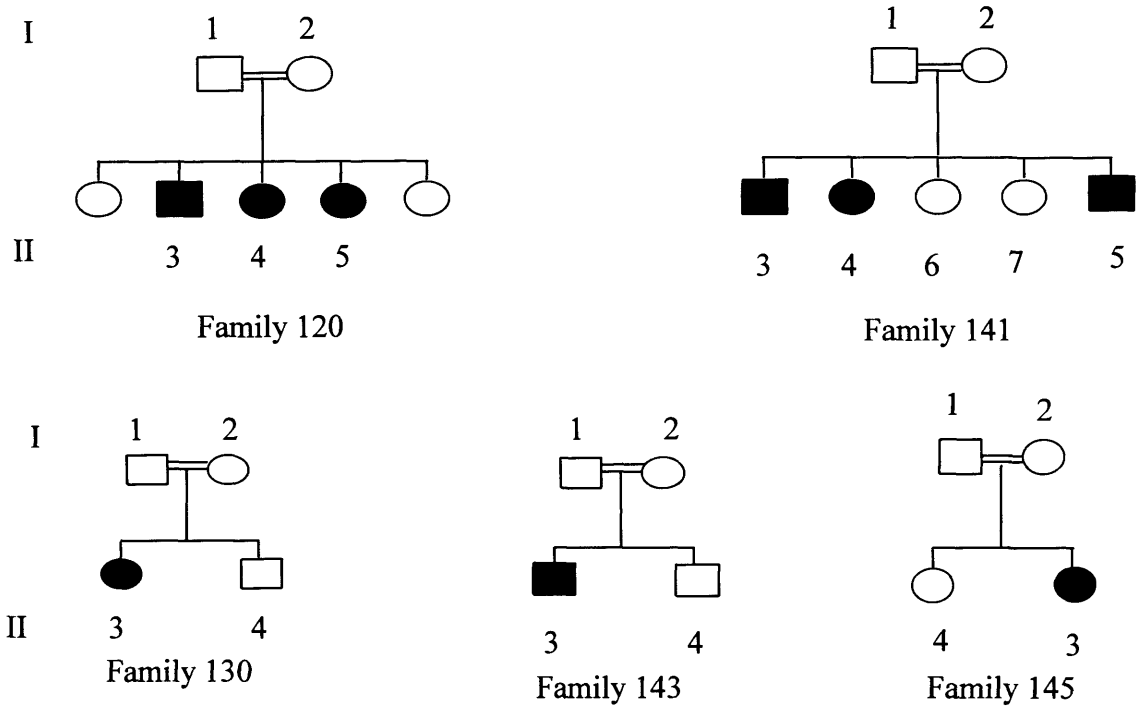


Figure 3.1: Pedigree structure of consanguineous families 120, 130, 141, 143 and 145. Only samples where DNA was available are numbered in figure. Family details are described more fully in methods section 2.3.

3.1.1 Power simulation for Pakistani families 120, 130, 141, 143 and 145

Prior to beginning the linkage study an analysis was made to evaluate whether the family resource available had the power to statistically prove linkage under a variety of assumed conditions and at the marker resolution being proposed. To this end, the five individual pedigrees (Figure 3.1) were analysed using the computer linkage programme SLINK (Lemire, 2006). This is a simulation programme that simulates marker data in pedigrees with respect to their disease locus phenotypes. The maximum LOD score that can be obtained under conditions of genetic homogeneity or of varying heterogeneity with the pedigrees of interest can be calculated.

The minimum and maximum achievable LOD scores for each Pakistani PCD family were calculated individually, assuming an autosomal recessive inheritance, a penetrance of 0.9 and a disease allele frequency of 0.007. Even with the same ethnic

origin and ciliary ultrastructural defect, genetic homogeneity is not guaranteed, therefore, in order to anticipate heterogeneity and take this into consideration within this group of families, simulations were also performed for combinations either of the two large family with three-affected individuals (family 120 or 141) plus all three smaller families with one affected individual (families 130, 143 and 145), as well as all five families together (Table 3.1). The maximum LOD scores that could be achieved directly at the disease locus ($\theta = 0$) for either of the larger families alone (family 120 or 141) was 2.4, whereas the single-affected families with (families 130 and 145) or without (family 143) one unaffected child (due to limited DNA availability) achieved 1.3 and 1.2, respectively. Since the LOD scores for more than one family are additive, if either of the larger three-affected families were linked to the same region as all the single-affected families, a maximum LOD score of 6.1 could be achieved. If all five families were linked to the same region, a maximum LOD score of 8.5 could be reached. This analysis indicated that these families would provide sufficient power in a genome-wide linkage screen to generate a significant LOD score of >3. Since the maximum LOD score achievable for family 120 or 141 is not sufficient to prove linkage, a significant LOD score of >3 could only be achieved if at least one three-affected and one single affected family were linked to the same region.

3.1.2 Whole genome SNP-based linkage analysis in Pakistani families

SNP genotyping using the *Illumina* SNP-based Linkage IVb Panel, as described in Section 2.3, was performed on DNA samples obtained from all nine affected individuals, seven (unaffected) parents (individuals 130 I:1, 145 I:1 and 145 I:2 were not included) and two unaffected siblings (individuals 141 II:6 and II:7 and 143 II:4 were not included) of the five Pakistani PCD families. Multipoint parametric linkage analysis was carried out on the SNP genotype data using the MERLIN programme (Abecasis et al., 2002), assuming an autosomal recessive inheritance, a disease allele frequency of 0.007 and a disease penetrance of 0.9.

All five families were included together in the MERLIN linkage analysis and significant HLOD scores of 3.7 (proportion of linked families, $\alpha = 0.8$) on

chromosome 11 and 3.9 ($\alpha = 0.7$) on chromosome 17 were achieved (Table 3.2). An HLOD score of 2.3 ($\alpha = 0.5$) was achieved on chromosome 9, which is described in more detail in Appendix 1. Linkage results for each family individually were examined and revealed that the larger families 141 and 120 contributed to the majority of the LOD score on chromosome 11 and 17, respectively, since they achieved the maximum simulated LOD score of 2.4 at these regions of interest (Table 3.3). In addition a region on chromosome 9 achieved a LOD score of 2.0 in family 120 (Appendix 1). No other genomic region associated with a LOD score reaching 2 was identified, indeed for the majority of chromosomes the maximum LOD score was below 1.0 (Appendix 1). Thus, locus heterogeneity was demonstrated as likely in these PCD families from the Mirpur region of Pakistan despite them originating from the same geographical region and having a similar ultrastructural defect observed by EM. Subsequent work focused on refining the loci on chromosomes 11 and 17 only.

	LOD scores															
	Family 120		Family 130		Family 141		Family 143		Family 145		Family 120 + 130,143,145		Family 141 + 130,143,145		All families	
θ	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
0	-3.186	**2.378**	-1.549	1.288	-5.076	**2.378**	-1.668	1.164	-1.549	1.288	-7.952	6.118	-9.842	6.118	-13.028	8.496
0.05	-0.917	2.156	-1.171	1.152	-0.928	2.156	-0.884	1.033	-1.171	1.152	-4.143	5.493	-4.154	5.493	-5.071	7.649
0.1	-0.76	1.928	-0.868	1.014	-0.787	1.931	-0.81	0.899	-0.868	1.014	-3.306	4.855	-3.333	4.858	-4.093	6.786

Table 3.1: SLINK simulation of minimum and maximum achievable LOD scores for the Pakistani families. LOD scores were calculated for each family individually, for combinations of either large three-affected family (120 or 141) plus three single affected families (130, 143, and 145) and for all five families together. Numbers in bold highlight the maximum LOD score achievable at $\theta = 0$ for the larger families.

Chromosome	All families HLODmax	Alpha
1	1.098	0.377
2	1.332	0.752
3	1.431	0.497
4	0.352	0.192
5	1.185	0.401
6	0.240	0.166
7	1.087	0.367
8	0.319	0.172
9	**2.334**	**0.530**
10	0.314	0.169
11	**3.731**	**0.764**
12	0.976	0.303
13	1.302	0.241
14	0.314	0.169
15	0.238	0.288
16	0.476	0.267
17	**3.868**	**0.738**
18	0.326	0.177
19	1.802	0.431
20	0.357	0.197
21	0.256	0.186
22	0.000	0.000

Table 3.2: Maximum HLOD scores calculated for the Pakistani families using *Illumina* SNP data and MERLIN linkage analysis on each chromosome. Maximum HLOD scores for all five Pakistani families with alpha values (proportion of linked families) are displayed.

3.1.3 Refinement of critical regions of interest in Pakistani families

MERLIN linkage analysis on the *Illumina* SNP genotype data of the five Pakistani PCD families 120, 130, 141, 143 and 145 identified regions of interest on chromosome 11 and 17 with HLOD scores >3 (α = 0.8). Observation of the SNP genotype data in affected individuals revealed homozygosity across these regions in the families. The largest region of shared homozygosity between the affected individuals within the region on chromosome 11 spanned 22.1 cM (11.6 Mb) between SNP markers *rs644252* and *rs2001625*. The largest region of shared homozygosity on chromosome 17 spanned 14 cM (16.4 Mb) between SNP markers *rs1859212* and *rs2045418*. In addition, a less significant region of homozygosity on chromosome 9 with an HLOD score of 2.3 (α = 0.5) was identified that spanned 12.1 cM (9.6 Mb) between SNP markers *rs1413219* and *rs769012*. The constructed haplotypes will be further described in section 3.1.4 and Appendix 1 for the three regions of interest.

In order to refine the limits of the critical regions, extra microsatellite markers were genotyped across the regions of interest on chromosomes 11 and 17, in the Pakistani families, including unaffected siblings where DNA was available. Eight additional known microsatellite markers were genotyped across the candidate region identified on chromosome 11. Eleven additional known and 3 in-house microsatellite markers (primer sequences shown in Appendix 1) were genotyped across the candidate region on chromosome 17.

MERLIN linkage analysis was repeated with the same parameters as used in section 3.1.2, using the more detailed genetic marker data incorporating microsatellite markers. From these results, the region on chromosome 11 achieved an HLOD score >3 across 3.3 Mb between markers *rs617847* and *rs6590098*. A maximum HLOD score of 3.6 (α = 0.8) was obtained on chromosome 11 at SNP marker *rs930983* (Table 3.3 and Figure 3.2A). The region on chromosome 17 achieved an HLOD score >3 between markers *D17S951* and *Ch17-nr-D17S1865* (8.8 Mb). A maximum HLOD score of 6.0 (α = 1.0) was obtained on chromosome 17 at in-house microsatellite marker *D17S809* (Table 3.4 and Figure 3.2B). The families with one affected individual 130, 143 and 145 each contributed a maximum LOD score of 0.6,

1.3 and 1.0, respectively to the locus on chromosome 11 and approximately 1.3 each to the locus on chromosome 17 (Tables 3.3 and 3.4).

		Illumina screen SNPs only		All 5 families - Illumina screen SNPs plus extra microsatellite markers			LOD scores for individual families - Illumina screen SNPs plus extra microsatellite markers				
Marker name	**Genetic distance (cM)**	**HLOD score**	**alpha**	**LOD**	**alpha**	**HLOD**	**Family 120**	**Family 130**	**Family 141**	**Family 143**	**Family 145**
rs644252	121.30	0.001	0.005	-8.505	0.000	0.000	-4.039	-0.151	-1.121	-2.282	-1.142
D11S4129	121.58			-8.738	0.000	0.000	-4.053	-0.626	2.076	-2.425	-1.144
D11S4132	121.90			-9.649	0.000	0.000	-4.092	-0.993	2.364	-2.484	-1.047
rs1073636	122.31	1.493	0.018	-5.693	0.247	1.175	-4.209	-0.053	2.366	-2.53	-1.044
D11S925	123.00			-5.189	0.338	1.868	-4.293	0.586	2.366	-2.842	-1.208
D11S4107	124.00			-3.386	0.337	1.887	-3.558	0.555	2.371	-1.847	-1.138
rs481703	124.46	1.666	0.381	-3.398	0.331	1.875	-3.47	0.531	2.371	-1.73	-1.338
rs665035	124.50	1.665	0.351	-3.434	0.330	1.873	-3.472	0.528	2.373	-1.729	-1.371
rs575030	124.51	1.7	0.349	-2.651	0.357	1.919	-3.474	0.527	2.374	-1.731	-0.584
rs596437	124.88	2.147	0.372	-1.881	0.522	2.464	-3.484	0.501	2.374	-2.152	0.643
rs617847	124.92	2.146	0.523	-2.418	0.522	2.478	-3.49	0.498	2.374	-2.703	0.665
rs676943	125.79	3.532	0.513	0.311	0.750	**3.425**	-4.101	0.428	2.374	0.488	0.885
D11S4167	126.00			0.396	0.753	**3.505**	-4.102	0.409	2.374	0.563	0.914
rs668183	126.08	3.622	0.763	0.408	0.754	**3.522**	-4.109	0.402	2.375	0.581	0.922
rs1894078	126.91	3.731	0.764	0.748	0.755	**3.641**	-3.888	0.316	2.375	0.712	0.997
rs930983	127.54	3.702	0.762	0.796	0.751	**3.647**	-3.838	0.236	2.375	0.783	1.003
rs485345	128.18	3.638	0.757	0.757	0.743	**3.597**	-3.811	0.137	2.375	0.815	1.005
rs1943467	128.70	3.565	0.749	0.554	0.732	**3.538**	-3.933	0.035	2.375	0.838	1.004
D11S1353	129.20			0.556	0.716	**3.461**	-3.817	-0.094	2.375	0.856	1.001
rs1148109	130.11	3.253	0.738	0.342	0.653	**3.444**	-3.815	-0.526	2.375	1.09	0.982
rs6590016	130.43	3.151	0.653	-0.118	0.609	**3.390**	-3.923	-0.924	2.375	1.149	0.969
rs2512161	130.51	3.123	0.609	-0.324	0.597	**3.371**	-3.931	-1.13	2.375	1.163	0.962
rs1893702	130.63	3.076	0.596	-1.352	0.577	**3.331**	-3.922	-2.165	2.375	1.182	0.941
rs2276189	131.05	3.008	0.576	-1.337	0.570	**3.225**	-3.804	-2.146	2.375	1.245	0.757
rs6590098	131.35	2.89	0.573	-1.889	0.513	2.809	-3.791	-2.161	2.375	1.284	0.166
D11S933	131.38			-3.298	0.393	2.493	-3.802	-2.165	2.375	1.29	-1.233
rs4936975	131.51	2.717	0.566	-3.324	0.393	2.493	-3.832	-2.163	2.375	1.29	-1.23
rs593290	131.64	2.149	0.551	-3.287	0.393	2.492	-3.791	-2.165	2.375	1.29	-1.231
rs3740892	131.80	2.149	0.387	-3.273	0.393	2.491	-3.781	-2.161	2.375	1.289	-1.23
rs4627097	131.97	2.199	0.387	-2.597	0.418	2.541	-3.733	-2.165	2.375	1.289	-0.598
rs1425842	132.11	2.219	0.412	-2.452	0.427	2.561	-3.714	-2.156	2.375	1.288	-0.479
rs528638	132.61	2.254	0.421	-2.426	0.442	2.590	-3.824	-2.163	2.375	1.283	-0.328
rs663714	133.00	2.265	0.437	-2.310	0.447	2.582	-3.78	-2.11	2.375	1.262	-0.288
rs586566	133.12	2.266	0.442	-2.050	0.447	2.574	-3.503	-2.121	2.375	1.253	-0.286
rs2156449	134.14	2.245	0.442	-1.675	0.432	2.428	-2.755	-2.287	2.373	1.118	-0.357
D11S4110	135.93			-5.277	0.202	1.538	-2.661	-2.07	2.373	-1.957	-1.2
rs1944819	135.95	2.149	0.434	-2.658	0.349	1.969	-2.662	-2.073	2.373	0.666	-1.2
rs949095	135.95	2.149	0.388	-2.658	0.349	1.969	-2.662	-2.073	2.372	0.666	-1.2
rs762827	137.93	2.146	0.387	-3.100	0.386	2.325	-2.845	-2.783	2.372	1.116	-1.189
rs1027323	139.46	2.151	0.389	-3.550	0.392	2.412	-3.477	-2.778	2.373	1.215	-1.106
rs923811	139.98	2.15	0.388	-3.063	0.391	2.422	-3.473	-2.235	2.374	1.23	-1.179
rs1043654	140.94	1.997	0.388	-3.006	0.392	2.280	-3.55	-1.963	2.219	1.246	-1.175
rs2001625	143.19	0.218	0.15	-7.271	0.158	0.277	-3.518	-2.126	-1.083	1.257	-1.168

Table 3.3: Genetic linkage analysis of Pakistani families on chromosome 11q. LOD, alpha and HLOD score shown for all five Pakistani PCD families before and after high resolution marker typing using *Illumina* SNP genotyping results and additional microsatellite markers. Microsatellites highlighted in bold. Genetic distance in centimorgans (cM) according to the *Illumina* genetic map constructed by deCODE. HLOD >3 highlighted in red.

		Illumina screen SNPs only		All 5 families - Illumina screen SNPs plus extra microsatellite markers			LOD scores for individual families - Illumina screen SNPs plus extra microsatellite markers				
Marker name	**Genetic distance (cM)**	**HLOD score**	**alpha**	**LOD**	**alpha**	**HLOD**	**Family 120**	**Family 130**	**Family 141**	**Family 143**	**Family 145**
rs1859212	66.10	1.32	0.507	-2.510	0.571	1.766	-1.28	1.096	-4.335	1.243	0.766
D17S250	67.63			-2.630	0.574	1.824	-1.258	1.106	-4.539	1.275	0.786
rs1008723	68.56	3.361	0.742	-2.407	0.581	1.878	-1.163	1.056	-4.463	1.283	0.88
rs1526601	69.27	3.479	0.74	-2.530	0.578	1.887	-1.268	0.962	-4.501	1.287	0.99
rs1468259	69.47	3.508	0.739	-2.622	0.574	1.869	-1.353	0.896	-4.493	1.288	1.04
D17S1787	70.08			-8.534	0.367	1.162	-4.304	-2.243	-4.537	1.29	1.26
rs1008753	70.08	3.58	0.739	-3.953	0.381	1.188	-3.852	-1.568	-1.083	1.29	1.26
rs13851	71.68	3.732	0.738	-2.157	0.591	1.925	-0.921	1.237	-4.464	1.29	0.701
rs739769	71.92	3.75	0.738	-2.451	0.57	1.74	-0.912	1.269	-4.54	1.29	0.442
D17S951	72.08			-5.360	0.379	1.2	-1.091	1.279	-4.541	1.29	-2.297
D17S1861	72.69			-0.232	0.644	**3.649**	2.374	1.289	-4.512	1.29	-0.673
rs744281	72.70	3.804	0.738	-0.231	0.644	**3.65**	2.374	1.289	-4.512	1.29	-0.672
D17S950	73.20			-0.357	0.64	**3.642**	2.375	1.29	-4.603	1.29	-0.709
D17S931	73.74			-3.900	0.592	**3.487**	2.375	1.291	-7.193	1.234	-1.607
rs736604	73.75	3.868	0.738	-2.442	0.711	**3.797**	2.375	1.291	-7.13	1.235	-0.213
rs2013383	73.97	3.868	0.738	-1.039	0.788	**4.828**	2.375	1.291	-6.929	1.287	0.937
D17S1859	73.98			-0.808	0.788	**4.844**	2.375	1.291	-6.714	1.29	0.95
rs2051821	74.32	3.866	0.738	0.113	0.788	**4.751**	2.375	1.291	-5.7	1.192	0.955
rs733920	74.32	3.866	0.738	0.113	0.788	**4.752**	2.375	1.291	-5.7	1.192	0.955
rs962272	74.45	3.864	0.738	0.216	0.788	**4.814**	2.375	1.291	-5.66	1.255	0.955
17-Mir-1	74.50			3.449	0.791	**4.854**	2.375	1.291	-2.462	1.29	0.955
D17S1868	74.53			3.848	0.796	**4.862**	2.375	1.291	-2.062	1.29	0.954
rs1063647	76.31	3.66	0.756	2.158	0.781	**4.509**	2.375	1.29	-3.397	1.29	0.6
D17S1869	77.00			0.204	0.595	**3.545**	2.375	1.29	-3.252	1.29	-1.499
D17S-SPAG9	77.90			3.464	0.756	**3.951**	2.375	1.29	-0.136	1.29	-1.355
D17S809	79.00			5.950	1	**5.951**	2.375	1.29	0.198	1.29	0.797
rs792786	79.53	2.934	0.95	5.732	1	**5.733**	2.375	1.29	0.204	1.29	0.573
rs1508966	79.87	3.796	1	5.614	1	**5.615**	2.375	1.29	0.206	1.29	0.453
Ch17-nr-D17S1865	79.28			2.152	0.803	2.382	-1.081	1.29	0.206	1.29	0.447
rs2045418	80.09	1.055	0.483	-1.238	0.471	0.879	-1.081	1.29	0.271	-2.16	0.442
Ch17-AC019315	80.40			-1.128	0.479	0.901	-1.082	1.289	0.311	-2.083	0.437

Table 3.4: Genetic linkage analysis of Pakistani families on chromosome 17q. LOD, alpha and HLOD score shown for all five Pakistani PCD families before and after high resolution marker typing using *Illumina* SNP genotyping results and additional microsatellite markers. Microsatellites highlighted in bold. Genetic distance in centimorgans (cM) according to the *Illumina* genetic map constructed by deCODE. HLOD >3 highlighted in red.

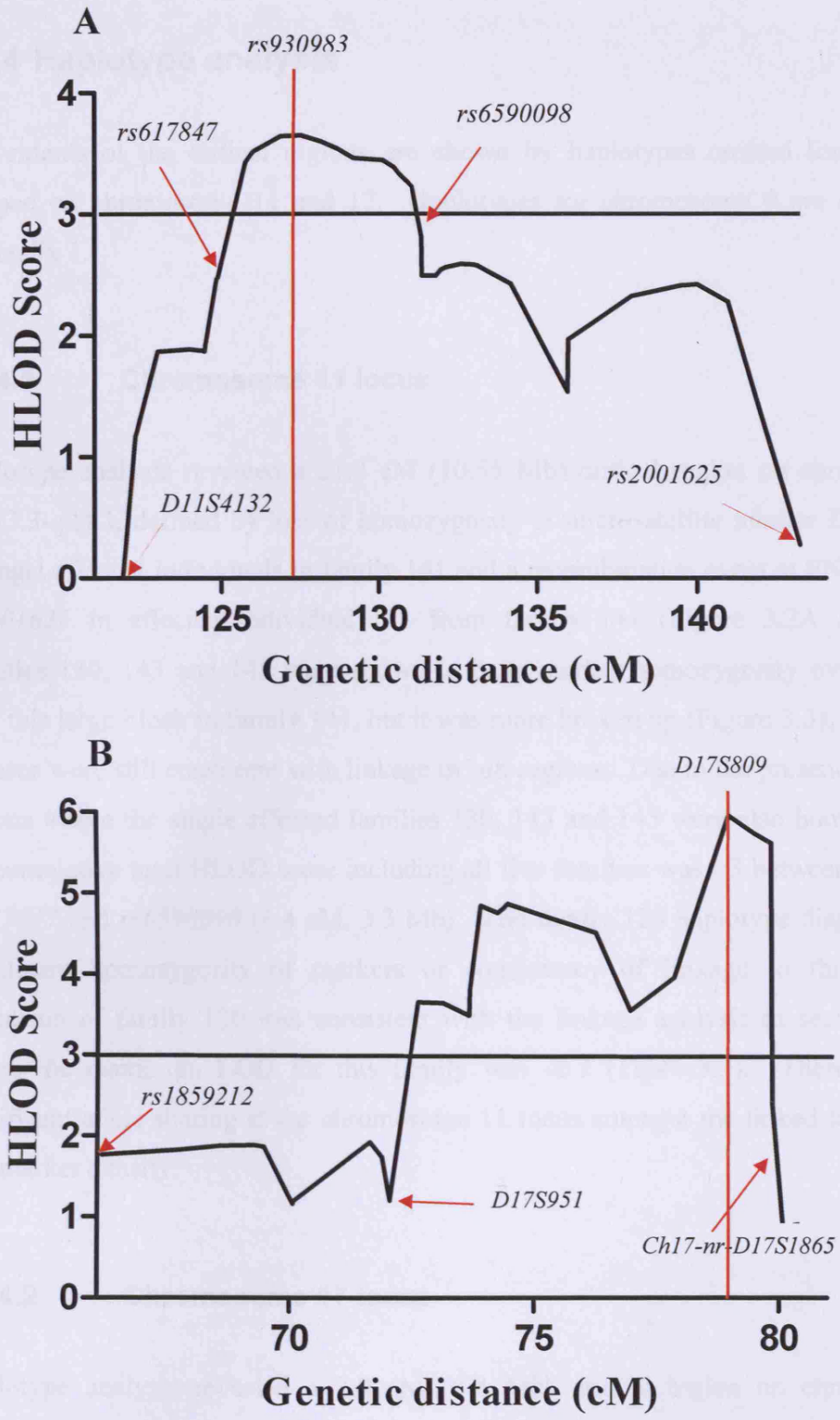


Figure 3.2: HLOD plot across region of interest on chromosome 11q and 17q for all five Pakistani PCD families. Genetic distance from the p arm telomere of the chromosome is plotted against HLOD score. A: Chromosome 11 HLOD plot. Maximum HLOD of 3.7 achieved at marker *rs930983* (124.32 cM). Critical region defined by loss of homozygosity at marker *D11S4132* amongst affected individuals in family 141 and a recombination event at marker *rs2001625* in affected individual II:3 from family 141. HLOD >3 between *rs617847* and *rs6590098*. B: Chromosome 17 HLOD plot. Maximum HLOD score of 6.0 achieved at marker *D17S809* (77.0 cM). HLOD >3 and critical region defined by recombination events in affected individual II:3 from family 120 at microsatellite markers *D17S951* and *Ch17-nr-D17S1865*, as shown in figure 4.

3.1.4 Haplotype analysis

The extents of the critical regions are shown by haplotypes created for the loci mapped on chromosome 11 and 17. Haplotypes for chromosome 9 are shown in Appendix 1.

3.1.4.1 Chromosome 11 locus

Haplotype analysis revealed a 21.3 cM (10.65 Mb) critical region on chromosome 11q23.3-q24.3, defined by loss of homozygosity at microsatellite marker *D11S4132* amongst affected individuals in family 141 and a recombination event at SNP marker *rs2001625* in affected individual II:3 from family 141 (Figure 3.2A and 3.3). Families 130, 143 and 145 were shown to have marker homozygosity overlapping with this large block in family 141, but it was more broken up (Figure 3.3), however, all three were still consistent with linkage in sub-regions. Due to the presence of sub-regions where the single affected families 130, 143 and 145 were also homozygous, the cumulative total HLOD score including all five families was >3 between markers *rs617847* and *rs6590098* (6.4 cM, 3.3 Mb). The family 120 haplotype displayed no significant homozygosity of markers or consistency of linkage to this region. Exclusion of family 120 was consistent with the linkage analysis in section 3.1.3 where the maximum LOD for this family was -2.7 (Table 3.3). There was no significant allele sharing at the chromosome 11 locus amongst the linked families at this marker density.

3.1.4.2 Chromosome 17 locus

Haplotype analysis revealed a 7.8 cM (8.8 Mb) critical region on chromosome 17q21.31-q22 defined by recombination events in affected individual II:3 from family 120 at microsatellite markers *D17S951* and *Ch17-nr-D17S1865* (Figure 3.2B and 3.4). The HLOD score at this locus was >3 due to overlapping homozygosity in affected individuals from families 130 and 143 along the length of the region and near-complete homozygosity in family 145. Affected individual II:3 from family 145 had a homozygous haplotype broken up by heterozygosity of three markers (Figure

3.4). The linkage analysis in section 3.1.3 suggested that all five families are genetically linked to this region (α = 1.0) between *D17S-SPAG9* and *Ch17-nr-D17S1865* (Table 3.4). However, the maximum LOD score of 0.3 in family 141 at this region was derived from only three homozygous markers *rs1508966*, *rs2045418* and *Ch17-ACO19315* at the telomeric end of the region, which were interrupted with heterozygous marker *Ch17-nr-D17S1865*, so family 141 was not considered significant for linkage. Again, there was no significant allele sharing at the chromosome 17 locus amongst the linked families at this marker density.

3.1.5 Heterogeneity amongst Pakistani families 120, 130, 141, 143 and 145

Haplotypes revealed that the positive LOD scores at the loci on chromosomes 11q and 17q derived mainly from either family 141 or 120, respectively. At both loci this larger family with three affected individuals that was consistent with linkage achieved a LOD score of 2.4. The true linkage status of the three singleton families 130, 143 and 145, was uncertain due to their insufficient linkage power.

The lack of allele sharing which would have indicated linkage disequilibrium suggested that if these families share a common mutation it is not detectable at this marker density. This was disappointing since the five families share a common ultrastructure and geographical origin.

3.1.6 Likelihood of linkage of Pakistani families to regions of interest

Both the regions mapped on chromosome 11 and 17 were consistent for linkage in a proportion of the Pakistani families. However, the haplotypes did not distinguish which locus was most likely for the singleton families. Formal heterogeneity testing using HOMOG3R (Bhat et al., 1999) was therefore performed to determine whether one locus was statistically more likely to contain the disease locus. HOMOG3R assumes more than one locus and works by calculating the log likelihood of linkage to a locus under the assumption that a trait is linked in a proportion of families in a

group to one marker within a locus, and in other families in the group it is linked to a different marker in another area of the genome.

HOMOG3R analysis confirmed that each of the two larger families 120 and 141 were consistent for linkage to only one of the two loci (Table 3.5). Family 120 had a 92% probability of linkage to chromosome 17q and family 141 had a 95% probability of linkage to chromosome 11q. Single affected families 130, 143 and 145 were shown to have approximately 50% probability of linkage to either locus (Table 3.5). Thus, it was most likely that family 120 or 141 linked to the disease locus at chromosome 17 and 11, respectively, but that the smaller families might be linked to either locus with almost equal probability. This heterogeneity test confirmed that four out of five families were consistent for linkage to each locus.

	Chromosome	
Family ID	**11q**	**17q**
120	0.0784	0.9216
130	0.4598	0.5402
141	0.9523	0.0477
143	0.4827	0.5173
145	0.5000	0.5000

Table 3.5: HOMOG3R analysis of the probability of linkage for loci identified in Pakistani families.

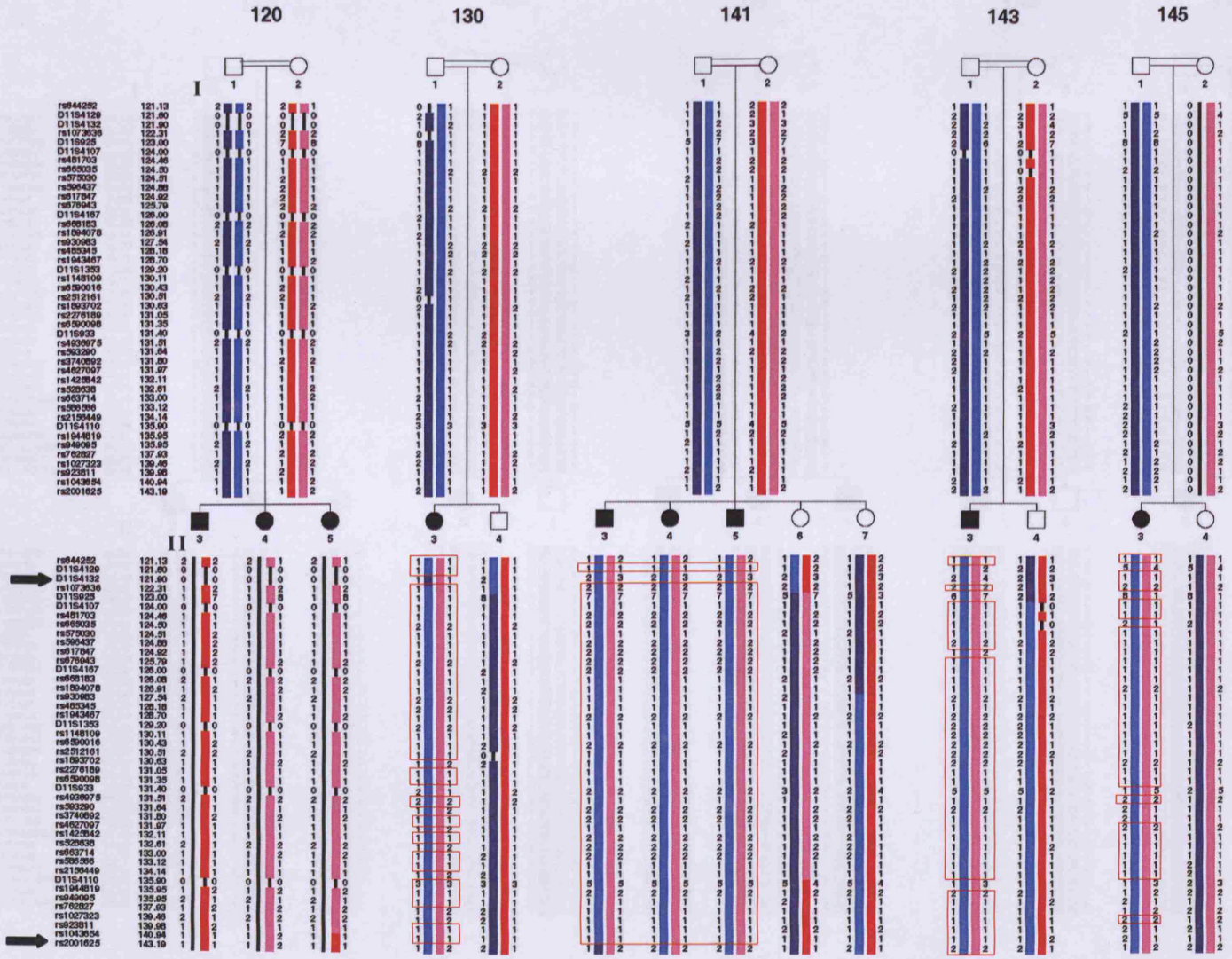


Figure 3.3: High density haplotyping of Pakistani families at the putative locus on chromosome 11q. Genetic distance from the tip of the p arm of chromosome 11 in cM (centimorgans). Boxing indicates homozygosity. Arrow indicates flanking markers. Note, Individuals 130 I:1, 145 I:1, 145 I:2, 141 II:6, 141 II:7 and 143 II:4 were not included in the *Illumina* SNP screen so their SNP haplotypes are inferred from microsatellite genotype information.

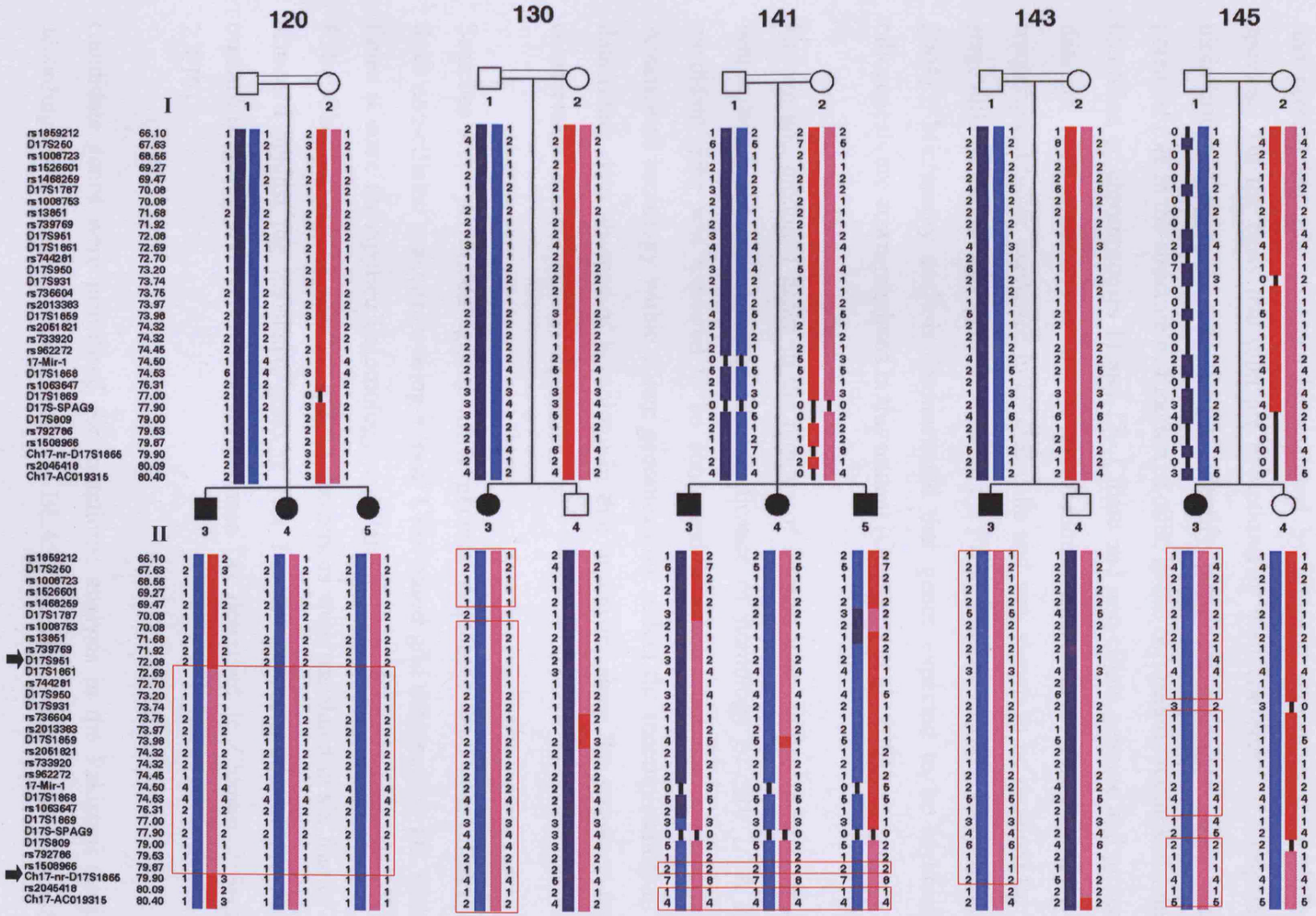


Figure 3.4: High density haplotyping of Pakistani families at the putative locus on chromosome 17q. Genetic distance from the tip of the p arm of chromosome 17 in cM (centimorgans). Boxing indicates homozygosity. Arrows indicate flanking markers. Note, Individuals 130 I:1, 145 I:1, 145 I:2 and 143 II:4 were not included in the *Illumina* SNP screen so their SNP haplotypes are inferred from microsatellite genotype information. Unaffected individuals from family 141 are not shown.

3.1.7 Identification of positional candidate genes at the Pakistani family loci on chromosome 11 and 17

During evolution, the genetic information of biological processes and organelles that are no longer required tends not to be preserved (Merchant et al., 2007). Essential pathways and components are conserved and this is reflected by conservation of gene and protein structure that can be analysed by aligning sequences from different species. On the basis that cilia are evolutionarily well conserved, a comparative bioinformatics approach was used to identify positional candidate genes with a potential role in the structure or function of cilia, in the two genetic regions of interest identified on chromosomes 11 and 17. Ciliate and non-ciliate genome and proteome datasets were used as well as the *Foxj1* mouse microarray. *Foxj1* is an f-box transcription factor expressed in ciliated cells and was shown to be required for late stage cilia formation by a gene deletion in the $Foxj1^{-/-}$ mouse model (Gomperts et al., 2004). Microarray analysis demonstrated that genes expected to be involved in ciliogenesis are down-regulated in this mutant (Gomperts et al., 2004).

Known and predicted genes in the regions of interest were subjected to a BLAST search using a cut-off E value for significance of homology of $\leq 10^{-5}$. A PCD candidate gene was expected to be conserved in ciliates but not in non-ciliates. Absence of homology within ciliate genomes may reflect the incompleteness of the data rather than absence of homology in that organism, since the sequence freeze completeness varied between organisms.

Searches were performed against whole genome sequences from nine ciliated and three non-ciliated organisms, using a local Unix-based grid designed by Dr. Richard Emes (Centre for Applied Entomology and Parasitology, Keele University) and Dr. Rahul Chodhari (UCL). Human mRNA sequences were translated in six frames and screened against the translated genome (six frames) or proteomes of the model organisms, using the TBLASTX programme (as described in Chapter 2 Section 2.4.10).

Candidate genes were prioritised for mutational analysis in the Pakistani families according to the results obtained from the BLAST search, in addition to information

A bioinformatics approach to create a priority list of genes, as descri
3.1.7, was performed on the chromosome 11q region of int
microsatellite markers *D11S4132* and *rs2001625*, the region whe
displays excess homozygosity. See Appendix 1 for full bioinformatic
the 74 genes in this region, 54 had a known function, 5 had a predicte
15 were hypothetical genes with no predicted function. For fifty-two g
possible to more definitively exclude them as candidates, as they had a
published non-ciliary function. After looking at the proposed f
remaining 22 genes, tissue expression, BLASTp homology and publi
two hypothetical genes were identified as good candidate PCD gen
LRRC35 (Table 3.6).

3.1.7.1.1 *SPA17* (Sperm Auto-antigenic protein 17)

SPA17 is located between markers *rs2276189* and *rs6590098* where th
was 2.8-3.2 (Figure 3.5A) and was considered a good positional c
because of its homology within the defined E value of $\leq 10^{-5}$ to gene
intestinalis genome and the *Tetrahymena thermophilia* genome. I
murine homologue is down-regulated in the cilia-less *Foxj1* knock-out
3.6).

The primary function of SPA17 is not clear but several published stu
role in cilia function. It was initially associated with binding to th
pellucida during fertilisation (Kong et al., 1995), however it has
become clear that SPA17 is likely to have a broader role as it has b
immunohistochemistry to be abundantly expressed in the cytoplasm
epithelial cells of the larynx, trachea, lung and fallopian tubes, as wel

section of the sperm flagellum (Grizzi et al., 2004). Expression in mouse spermatozoa and testis has been shown by western blot, immunoprecipitation and immunolocalisation (Grizzi et al., 2004; Frayne and Hall, 2002). RT-PCR also demonstrated *SPA17* expression in a wide range of human tissues although at a lower level than in testis (Frayne and Hall, 2002). SPA17 was shown to co-localise with acetylated tubulin in the cilia layer of the olfactory epithelium (McClintock et al., 2008). Although it has not yet been localised to the axoneme, a role in cilia and flagella could be postulated for SPA17 due to its expression pattern and localisation to cilia and flagella. The N-terminus of SPA17 was shown by sequence analysis to contain a 74 bp PKA regulatory subunit II (RII) domain, which contains the dimerisation domain that is responsible for the interaction between two individual RII subunits, required for binding A-kinase anchoring proteins (AKAPs) (Frayne and Hall, 2002). AKAPs function as scaffolding molecules that co-ordinate the actions of several signalling proteins and the RII domain of AKAP3 has been shown to bind to SPA17 along the length of the sperm flagella (Lea et al., 2004), thus SPA17 could be suggested to have a scaffolding/signalling role within cilia and flagella, i.e. in the regulation of dynein arm activity and the control of ciliary beat frequency.

The NCBI BLASTp search revealed homology of SPA17 to *Chlamydomonas* axonemal radial spoke proteins RSP11 and RSP7, however this was mostly due to non-specific homology based on the RII domain of the regulatory PKA subunit portion. Although the exact role of SPA17 in cilia function is unclear, evidence supporting its prioritisation as a good positional candidate relates to its specific localisation in cilia and flagella and putative scaffolding/signalling role.

3.1.7.1.2 *LRRC35* (Leucine Rich Repeat-containing protein 35)

LRRC35 is located between markers *D11S4107* and *rs481703* where the HLOD score was 1.9 (Figure 3.5A), just on the telomeric edge of the critical region but it was considered a good positional candidate due to homology to other potential axonemal genes, which is discussed later in this section. Bioinformatic screening initially revealed that homology was present to genes in all the datasets, nine ciliated and three non-ciliated, except for the *Trypanosoma* flagella proteome and the library of genes down-regulated in the *Foxj1* knockout mouse (Table 3.6). Expression in non-ciliates

was not expected for PCD genes, however, it is possible that the homology observed in the bioinformatic screen was due to a large proportion of the protein containing a leucine rich repeat (LRR), which forms part of a large protein family. In fact, published literature specified that *LRRC35* is not expressed in *Arabidopsis* or *S. cerevisiae* and *S. pombe* (Bartolini et al., 2005) and this identified that that BLAST homology was indeed non-specific. *LRRC35* is expressed at a particularly high level in testis, but is also present in brain, kidney, liver and lung at a lower level (Bartolini et al., 2005) (Table 3.7).

An NCBI BLASTp revealed that *LLRC35* had significant homology to tubulin folding cofactor E (*TBCE*). *LRRC35* has also been termed 'TBCE-like' (Bartolini et al., 2005). It is thought to function as a regulator of tubulin stability (Bartolini et al., 2005).

Two LRR genes containing six leucine rich repeats (LRR) were identified as positional candidate genes in this chapter: *LRRC35* within the chromosome 11 locus and *LRRC46* within the chromosome 17 locus (Section 3.1.7.2.2). The LRR domain creates a curved surface on the protein formed with parallel β-sheets linked by α-helices and functions in subunit-subunit interactions (Freshour et al., 2007). Both candidate genes *LRRC35* and *LRRC46* are similar to human *LRRC50* due to high homology within the LRR region and also conservation of residues at the C-terminus (Figure 3.6). *LRRC50* is of interest as it is the human orthologue of *Chlamydomonas ODA7*. Outer-Inner-Dynein (OID) linkers are novel axoneme structures that have recently been revealed by cryoelectron tomography that allows higher resolution than previously achievable (Nicastro et al., 2006). These are thought to coordinate inner and outer dynein arm activities (Freshour et al., 2007). In *Chlamydomonas* the ODA7 protein interacts with both the inner and outer dynein arms providing a structural OID link (Freshour et al., 2007). In addition, zebrafish mutant *hu255H* was found to have a mutation within the *LRRC50* orthologue, *lrrc50*, and mutants have ciliary defects affecting dynein arms and misalignment of peripheral microtubules. This results in phenotypes consistent with cilia defects in zebrafish including the development of pronephric cysts and abnormal left-right axis determination (van et al., 2008). Human LRRC35 is 26 % and LRRC46 is 30 % identical to *Chlamydomonas* ODA7 (Figure 3.6).

Candidate genes *LRRC35* and *LRRC46* also showed homology to an outer dynein arm (ODA) component, axonemal dynein light chain 1 (LC1) (Figure 3.7). Human *LRRC50* diverged from axonemal dynein LC1 (Freshour et al., 2007) and the LRRC35 and LRRC46 proteins are mainly similar at the LRR repeats, less so at the N- and C-terminus. The identity between human LC1 and LRRC35 is 21 % and LRRC46 is 29 %. Thus, it is important to note that LC1 and LRRC50 are different proteins but are similar with respect to the LRR domains.

Homology to axonemal *LC1* and *LRRC50/ODA7* made *LRRC35* and *LRRC46* excellent candidates for PCD. However, it is useful to carefully examine the homology alignments to understand if these two proteins could have an orthologous role in the human axoneme or whether there is more spurious homology based simply on the LRR domain. Homology of both *LRRC35* and *LRRC46* was greater to *LRRC50* than *LC1*, with *LRRC46* having higher homology than *LRRC35* to *LRRC50*. In conclusion, the majority of the homology derived from the LRR domain but some conservation of residues outside of this domain as well.

3.1.7.2 Candidate genes at the chromosome 17 locus consistent with linkage in Pakistani families 120, 130, 143 and 145

The same bioinformatic approach was applied to the chromosome 17q region of interest where family 120 displayed homozygosity, between microsatellite markers *D17S951 and Ch17–near D17S1865* (Figure 3.5B). See Appendix 1 for full bioinformatic screen. Out of the 139 genes in this region, 116 had a known or predicted function and 23 an unknown function. Out of those with a known or predicted function, 115 were excluded as positional candidates because of their predicted non-ciliary function. One gene, annotated with a predicted function, *FLJ35808*, was considered a candidate gene of interest due to its predicted tubulin tyrosine ligase homology and tissue expression. Of the 23 genes within the region with unknown function, one was considered a good candidate: *LRRC46*, because of its tissue expression and significant homology to genes with a role in the structure and/or function of flagella (Table 3.6 and 3.7).

3.1.7.2.1 *FLJ35808* (Hypothetical protein)

FLJ35808 is located on chromosome 17 between markers *D17S950* and *D17S931* where the HLOD score for this group of families was 3.6-3.5 (Figure 3.5B). It was considered to be one of the good positional candidates because cross-species conservation for the encoded protein was not observed in the three non-ciliate species screened but conservation was present in the *Chlamydomonas reinhardtii* genome, the *Trypanosoma brucei* and *Leishmania major* combined genomes, the *Ciona intestinalis* genome, the *Tetrahymena thermophilia* genome, the *C. elegans* proteome (Table 3.6). The murine homologue was also down-regulated in the *Foxj1* mouse. Expression profiling with reference to NCBI expressed sequence tag (EST) counts shows specific expression in human brain, lung, trachea and testis (Unigene – http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) (Table 3.7).

The gene was annotated in NCBI as having tubulin tyrosine ligase activity, catalysing the ATP-dependent post-translational addition of tyrosine to the C-terminal end of α-tubulin (Janke et al., 2005). This is based on derived homology rather than published experimental evidence. The *in vivo* consequences of this tyrosination activity is not yet established, although it has been shown that the tyrosination of the C-terminus of β-tubulin is required for ciliary beating (Vent et al., 2005). It is thought that a reversible tyrosination activity of α-tubulin could affect microtubule organisation and distribution of microtubule associated proteins (MAPs) (Vent et al., 2005). Thus, mutations in this candidate gene could cause a disruption of ciliary microtubules and give rise to PCD.

3.1.7.2.2 *LRRC46* (Leucine Rich Repeat containing protein 46)

LRRC46 is located on chromosome 17 between markers *D17S950* and *D17S931* where the HLOD score for this group of families was 3.6-3.5 (Figure 3.5B). It was considered a good positional candidate at chromosome 17q because it had homology within the cut-off E-value $\leq 10^{-5}$ to a gene or protein in all nine ciliate databases screened (Table 3.6). However, conservation was also observed in two non-ciliate datasets, *S. cerevisiae* and *S. pombe*. *LRRC46* is expressed in lung, trachea and testis according to the UniGene expression profile analysis of EST counts (Table 3.7).

As discussed in section 3.1.7.1.2. *LRRC46* had homology to LRR protein *LRRC50*, the human orthologue of *Chlamydomonas ODA7*, an axonemal OID linker protein and *Chlamydomonas* axonemal dynein LC1 (Figure 3.7). Human LRRC46 is 30 % identical to *Chlamydomonas* ODA7 (Figure 3.6). Thus, of the two positional candidate genes within the locus refined on chromosome 17q, *LRRC46* was considered the most promising candidate for carrying a potential disease-causing mutation as it is possible that mutations in the *LRRC46* gene could cause disruption of the I/ODA complex, consistent with the Pakistani family ultrastructural phenotype of missing inner and outer dynein arms, and thereby cause PCD.

Gene	Chromosome	ARAB	CERV	POMB	CHLY	PAZO	DEFLA	TRYP	CION	TETR	McKE	ELEG	FOXJ1
Sperm autoantigenic protein 17 (*SPA17*)	11q	0	0	0	0	0	0	0	1	1	0	0	1
Leucine rich repeat containing 35 (*LRRC35*)	11q	1	1	1	1	1	1	1	1	1	0	1	1
Hypothetical protein FLJ35808 (*FLJ35808*)	17q	0	0	0	1	0	0	1	1	1	0	1	1
Leucine rich repeat containing 46 (*LRRC46*)	17q	0	1	1	1	1	1	1	1	1	1	1	1

Table 3.6: Comparative BLAST analysis of candidate genes within chromosome 11 and 17 loci consistent with linkage in the Pakistani family group.

Key to genome/proteome data:
ARAB - *Arabidopsis thaliana*
CERV - *Saccharomyces cerevisae*
POMB - *Saccharomyces pombe*
CHLY - *Chlamydomonas reinhardtii* genome
PAZO - *Chlamydomonas* Flagellar proteome
DEFLA - *Chlamydomonas reinhardtii* deflagellation dataset
TRYP - *Trypanosoma brucei* and *Leishmania major* combined genomes
CION - *Ciona intestinalis* genome
TETR - *Tetrahymena thermophilia* genome
McKE - *Trypanosoma* flagellar proteome
ELEG - *Caenorhabditis elegans* proteome
FOXJ1 – Genes down regulated in *Foxj1* cilia-less mouse

0 = no hit in the BLAST search 1 = hit in the BLAST search (Cut-off E value ≤10-5 – see Methods Section 2.4.10)

Gene	Chromosome	Function	Domains identified in Human NCBI BLASTp search	Homology	Tissue expression
Sperm autoantigenic protein 17 (*SPA17*)	11q	Possible multifunctional role in zona pellucida binding and a scaffolding/signalling role within cilia and flagella (P)	PKA regulatory subunit II domain (RII) (P)		Widespread, including brain, eye, kidney, liver, lung, testis, trachea (NCBI EST counts)
Leucine rich repeat containing 35 (*LRRC35*)	11q	Regulator of tubulin stability (A)	Leucine-rich repeat (LRR), homology to axonemal light chain 1 (LC1) (A)	Human axonemal dynein light chain 1 (LC1) (B), Tubulin folding cofactor E (TBCE) (P)	Widespread including brain, eye, kidney, liver, lung, trachea. Highly expressed in testis (P) (Bartolini)
Hypothetical protein FLJ35808 (*FLJ35808*)	17q	Tubulin tyrosine ligase activity (A)	Tubulin-tyrosine ligase family domain (A)	Tubulin Tyrosine Ligase (A)	Brain, lung, testis, trachea, intestine, stomach (NCBI EST counts)
Leucine rich repeat containing 46 (*LRRC46*)	17q	Unknown	Leucine-rich repeat (LRR) (A)	Outer dynein arm 7 (ODA7/LRRC50) (B), Human axonemal dynein light chain 1 (LC1) (B)	Brain, eye, lung, testis, trachea (NCBI EST counts)

Table 3.7: Function, domain structure and tissue expression of four positional candidate genes for the Pakistani family group. (P) – published experimental evidence (see text for details), (A) – annotation in genomic databases, (B) – in-house BLAST search.

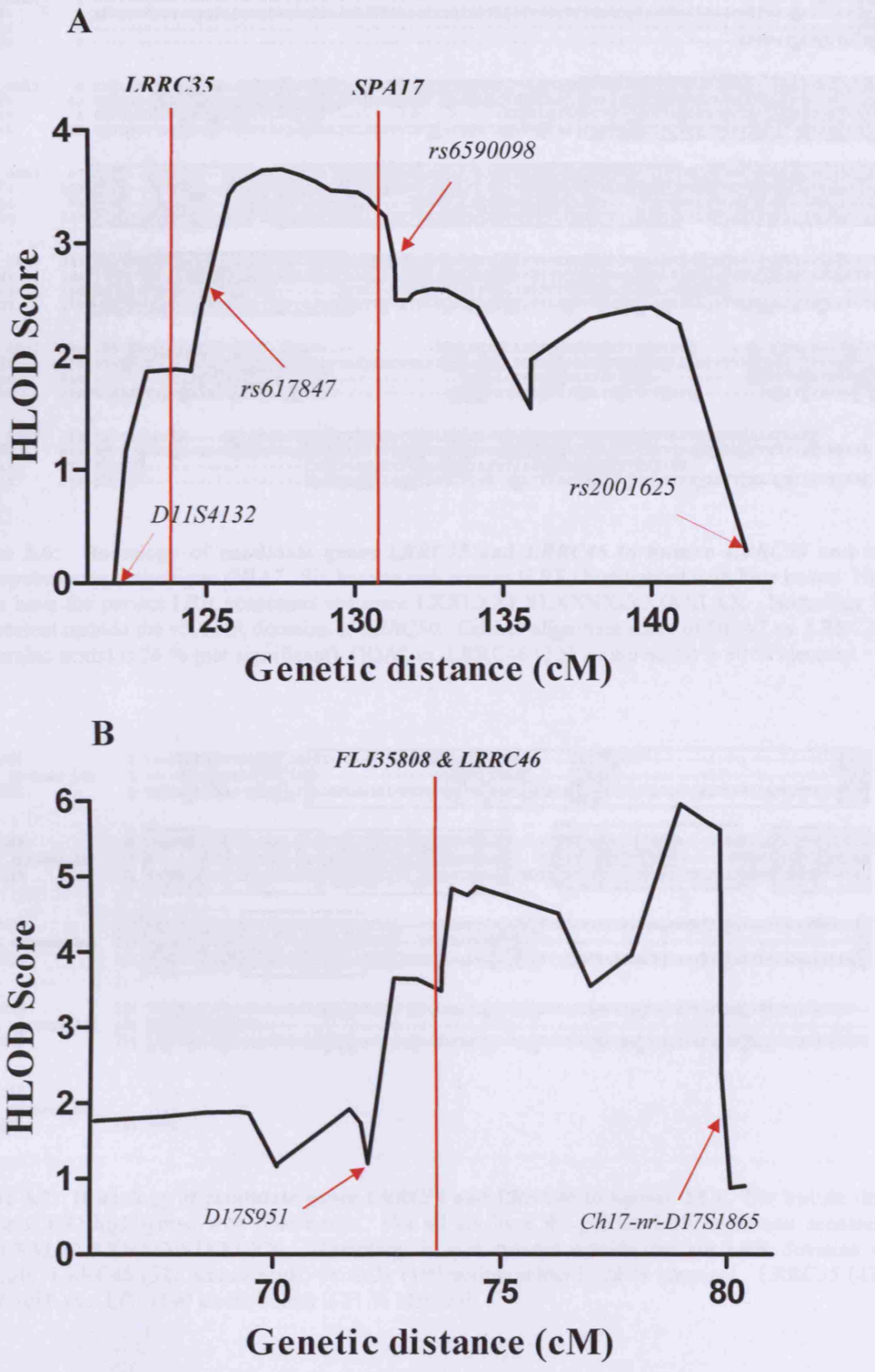


Figure 3.5: HLOD plots for chromosome 11q and 17q loci with position of candidate genes shown. A: Chromosome 11 candidate genes *LRRC35* and *SPA17*. B: Chromosome 17 candidate genes *FLJ35808* and *LRRC46*.

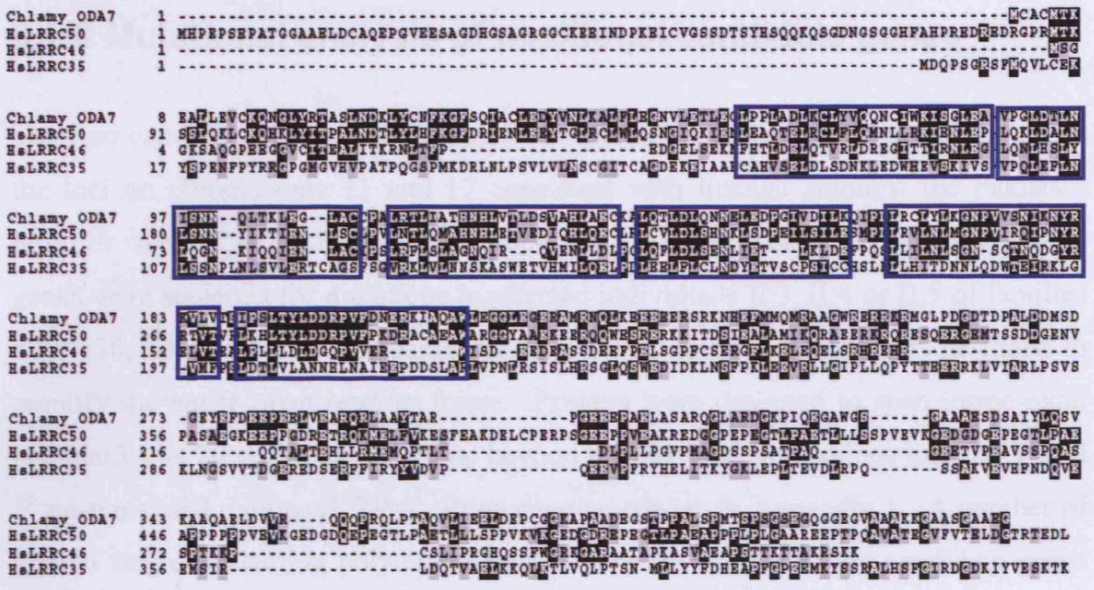


Figure 3.6: Homology of candidate genes *LRRC35* and *LRRC46* to human *LRRC50* and its *Chlamydomonas* orthologue *ODA7*. Six leucine rich repeats (LRR) highlighted with blue boxes. Not all six have the perfect LRR consensus sequence LXXLXXLXLXXNXIXXIXXLXX. Homology is also present outside the six LRR domains of *LRRC50*. Critical alignment score of ODA7 vs. LRRC35 (424 amino acids) is 26 % (not significant), ODA7 vs. LRRC46 (321 amino acids) is 30 % identical.

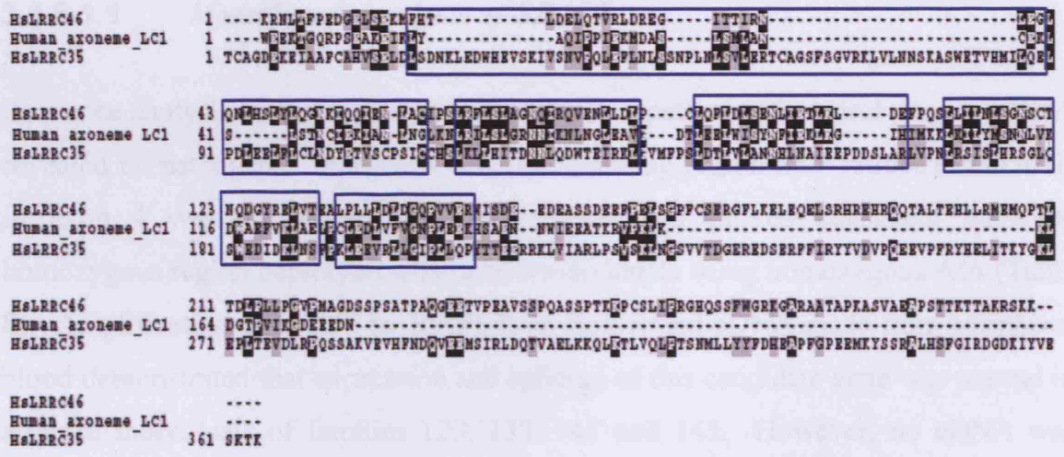


Figure 3.7: Homology of candidate genes *LRRC35* and *LRRC46* to human LC1. Six leucine rich repeats (LRR) highlighted with blue boxes. Not all six have the perfect LRR consensus sequence LXXLXXLXLXXNXIXXIXXLXX. Homology is also present outside the six LRR domains of *LRRC50*. LRRC46 (321 amino acids) vs. LC1 (190 amino acids) is 29 % identical. LRRC35 (424 amino acids) vs. LC1 (190 amino acids) is 21 % identical.

3.1.8 Mutational analysis of positional candidate genes

The four candidate genes identified and considered good positional candidates within the loci on chromosome 11 and 17 consistent with linkage amongst the Pakistani families were *SPA17, LRRC35, FLJ35808* and *LRRC46* (Table 3.7). The candidate genes were screened for mutations in affected individuals II:3, II:4 or II:5 of families 120, 130, 141, 143 and 145, using genomic DNA samples and primers designed to amplify the entire open reading frame. Primers were designed to span intron-exon boundaries by 20 bp, or extend 40 bp beyond the start and stop codons into the 5' and 3' un-translated regions (UTRs). Primer sequences are in Appendix 1. A number of known single nucleotide polymorphisms (SNPs) are located in the candidate genes (Table 3.8).

3.1.8.1 Sequencing of Chromosome 11q Candidate Genes

3.1.8.1.1 *Mutational Analysis of SPA17*

Sequence analysis of the candidate gene *SPA17* in patient and control genomic DNA revealed no pathogenic mutations, although a coding region SNP *rs26001463* (A/G) in exon 4 was polymorphic in the families and proved consistent with the homozygous region haplotype, all affected individuals being homozygous A/A (Table 8). Amplification of *SPA17* by RT-PCR on first-strand cDNA made from peripheral blood demonstrated that expression and splicing of this candidate gene was normal in affected individuals of families 120, 130, 143 and 145. However, no cDNA was available for the main linked family 141.

3.1.8.1.2 *Mutational Analysis of LRRC35*

Sequence analysis of candidate gene *LRRC35* in patient and control genomic DNA extracted from blood of affected individuals revealed no pathogenic mutations or polymorphic SNPs within the coding regions.

3.1.8.2 Sequencing of chromosome 17 candidate genes

3.1.8.2.1 *Mutational Analysis of FLJ35808*

Sequence analysis of candidate gene *FLJ35808* revealed no pathogenic mutations in DNA from affected individuals from all five families, although homozygous coding SNPs *rs2032844* and *rs8080242* (C/A and A/C) in exon 7 were polymorphic in the Pakistani families and proved consistent with the homozygous haplotype, all affected individuals being homozygous C/C or A/A, respectively (Table 3.8). Attempts to amplify *FLJ35808* by nested RT-PCR on first-strand cDNA made from peripheral blood revealed this gene was not expressed in high enough levels in blood to be able to amplify a product, even by nested PCR, and as a result this gene was not screened for mutations at the transcript level.

3.1.8.2.2 *Mutational Analysis of LRRC46*

Sequence analysis of candidate gene *LRRC46* in patient and control genomic DNA revealed no pathogenic mutations or polymorphic SNPs within the coding regions.

Gene	Chromosome	Number of coding exons	Coding SNPs
Sperm autoantigenic protein 17 (*SPA17*)	11q	4	Exon 4 A/G non-synonymous *rs26001463*
Leucine rich repeat containing 35 (*LRRC35*)	11q	7	No coding SNPs
Hypothetical protein FLJ35808 (*FLJ35808*)	17q	8	Exon 7 C/A non synonymous *rs2032844*, A/C synonymous *rs8080242*
Leucine rich repeat containing 46 (*LRRC46*)	17q	8	No coding SNPs

Table 3.8: Coding SNP information for the Pakistani positional candidate genes

3.2 *Discussion*

The aim of this study was to map a new PCD gene using homozygosity mapping and linkage analysis. Five PCD families 120, 130, 141, 143 and 145 (Figure 3.1) were grouped together in this chapter to increase the chance of genetic homogeneity because they all originated from the same geographical origin: Mirpur, the North West Frontier or the Rawalpindi region of Pakistan, before following a typical migration route to Bradford, UK. In addition, all nine affected individuals shared a common ultrastructural phenotype revealed by electron microscopy (EM) of missing inner and outer dynein arms and high speed video imaging revealed that they had a common motility defect of immotile or slightly flickering cilia.

All five families were included together in the MERLIN multipoint parametric linkage analysis on SNP genotype data generated using the *Illumina* SNP-based Linkage IVb Panel. Two loci were identified with significant HLOD scores >3 on chromosome 11 and chromosome 17 which were refined by genotyping of extra microsatellite markers, and haplotypes were analysed for homozygosity.

From the *Illumina* data, a maximum HLOD score of 3.7 was achieved at the region on chromosome 11 (proportion of linked families, $\alpha = 0.8$) across 22.1 cM (11.6 Mb) and initial observation of the SNP data revealed homozygosity in affected individuals between SNP markers *rs644252* and *rs2001625*. Additional microsatellite markers were genotyped across the region and MERLIN linkage analysis was repeated on each individual family using the more detailed genetic marker data incorporating microsatellites. From these results, the region with an HLOD score >3 was refined to 3.3 Mb between markers *rs617847* and *rs6590098*, with a maximum HLOD score of 3.7 ($\alpha = 0.8$) at SNP marker *rs930983*. The larger family 141 with three affected offspring achieved the maximum possible simulated LOD score of 2.4 at this region of interest and the families with one affected individual 130, 143 and 145 each contributed maximum LOD scores at different markers of 0.2, 0.8 and 1.0, respectively. Haplotype analysis of the locus on chromosome 11 revealed a 21.3 cM (10.65 Mb) critical region on chromosome 11q23.3-q24.3 amongst affected individuals in family 141, defined by loss of homozygosity at microsatellite marker *D11S4132* amongst affected individuals in family 141 and a recombination event at SNP marker *rs2001625* in affected individual II:3 from family 141. Families 130,

143 and 145 were shown to have intermittent marker homozygosity overlapping with this large block in family 141, which was consistent with linkage in sub-regions. The family 120 haplotype displayed no significant homozygosity of markers or consistency of linkage to this region.

From the initial *Illumina* data, a maximum HLOD score was achieved at the region on chromosome 17 (proportion of linked families, $\alpha = 0.8$) spanned 14 cM (16.4 Mb) between SNP markers *rs1859212* and *rs2045418* and initial observation of the SNP data revealed homozygosity in affected individuals between SNP markers *rs1859212* and *rs2045418*. Additional microsatellite markers were genotyped across the region and MERLIN linkage analysis was repeated on each individual family using the more detailed genetic marker data incorporating microsatellites. From these results, the region was refined to 8.8 Mb with an HLOD score >3 between markers *D17S951* and *Ch17-nr-D17S1865*. A maximum HLOD score of 6.0 ($\alpha = 1.0$) was obtained on chromosome 17 at in-house microsatellite marker *D17S809*. The large family 120 with three affected offspring contributed the maximum possible simulated LOD score of 2.4 for this family across this region of interest and the families 130, 143 and 145 with only one affected child contributed 1.3 each at different markers. Haplotype analysis of the locus on chromosome 17 was consistent with the linkage analysis and revealed a 7.8 cM (8.8 Mb) critical region on chromosome 17q21.31-q22 defined by recombination events in affected individual II:3 from family 120 at microsatellite markers *D17S951* and *Ch17-nr-D17S1865*. Families 130 and 143 displayed overlapping homozygosity across this region and family 145 had near-complete homozygosity here. The family 141 haplotype displayed no significant homozygosity of markers or consistency of linkage to this region.

Observation of the HLOD plot for the locus on chromosome 17 revealed some inconsistency in the linkage:

1. The HLOD score increased at marker *D17S-SPAG9* from 4.0 to 6.0 (Table 3.4, red numbers and Figure 3.5A) because the LOD scores for families 141 and 145 became more positive here, whereas the LOD scores for the other families 120, 130 and 145 were more uniformly consistent across the region.

Therefore, this inconsistency was mainly due to change in the LOD scores for families 141 and 145.

2. Across the three markers *D17S809*, *rs792786* and *rs1508966* the proportion of linked families, α, becomes 1.0, (Table 3.4, column 6) indicating that across this small 0.87 cM (2 Mb) region the multipoint linkage analysis assumed that all five families were consistent for linkage. This was because SNP marker *rs1508966* was observed by haplotype analysis to be fully informative and homozygous by descent (HBD) in affected individuals of family 141, which increased the multipoint LOD score at this region. However, this result is inconsistent since family 141 was excluded for linkage to the locus on chromosome 17 due to a recombination event in individual 141 II:3 at marker *D17S1869* and heterozygosity at all other markers across this region. In addition at this region, the lack of parental information for SNP markers *rs792786* and *rs1508966* in family 145 resulted in a more positive LOD score.
3. The HLOD score decreased to 3.5 across markers *D17S1869* and *D17S-SPAG9* (Table 3.4, red numbers) due to heterozygosity in affected individual 145 II:3, which resulted in a negative LOD score at these two markers for this family. Since this study was completed, these genotyping results were revisited and a thorough check of the raw data made. For microsatellite markers *D17S1869* and *D17S-SPAG9* the results were discovered to be unclear and these two markers had only been analysed once. Thus it is possible that they are not truly heterozygous.

To minimise locus heterogeneity, the five Pakistani families had been grouped together based on geographical origin and cilia ultrastructure. However, two loci were identified on chromosome 11 and 17 which were indistinguishable and the following conclusions can be drawn:

1. Although the maximum HLOD score at the locus on chromosome 17 was higher than on chromosome 11, haplotype analysis revealed that this was likely to be misleading due to marker uninformativity and recombination events in family 141.
2. Based on linkage and haplotype analysis it was clear that the positive HLOD scores on chromosome 11 and 17 were derived mainly from family 141 or

120, respectively, where a maximum LOD score of 2.4 were achieved in both cases.

3. The true linkage status of the three smaller families 130, 143 and 145 was uncertain due to insufficient linkage power and there was no obvious difference in homozygosity between the two regions.
4. HOMOG3R analysis was consistent with the haplotype analysis in that family 120 was linked to the region on chromosome 17 and family 141 was most likely to be linked to chromosome 17, whereas the smaller families 130, 143 and 145 were judged equally likely to be linked to either region.

Therefore, despite a common geographical origin and a common ultrastructural phenotype, genetic homogeneity is not guaranteed and heterogeneity still remains an obstacle in the mapping of PCD genes. In order to distinguish the two loci identified in this study further, more families would be required for genetic linkage analysis to these regions or to identify linkage disequilibrium (LD) to these regions, more marker typing could be performed. The lack of allele sharing amongst the linked families could indicate that only the two large families 120 and 141 are genuinely linked to the loci at chromosome 17 and 11, respectively, but that the homozygosity observed for the smaller families 130, 143 and 145 at these regions is spurious. It could also be possible that the families map to the same gene but harbour different mutations. Furthermore, it is possible that the LD region was smaller than the SNP density used of 600 kb and it was therefore missed in this genome screen. However, the LOD score of 2.4 on chromosome 17 and 11 for families 120 and 141 individually was the maximum achieved in the genome screen, with no other region achieving a LOD score >0.3, except for 2.0 for family 120 on chromosome 9 (Appendix 1). In the event of not finding any mutations in positional candidate genes on chromosome 11 or 17, chromosome 9 should be further investigated for family 120.

A comparative bioinformatic approach was used to identify positional candidates within the regions of interest based on the hypothesis that ciliary genes have been lost throughout evolution in non-ciliated organisms. Identifying those genes conserved in ciliated organisms and according to any putative functions related to cilia structure and/or function and their tissue expression allowed prioritization of 4 out of the >200 positional candidate genes within the two regions on chromosome 11 and 17:

FLJ35808, *LRRC35*, *SPA17* and *LRRC46*. DNA sequencing of the coding regions of these candidate genes in the linked families did not reveal any mutations, although coding SNP analysis demonstrated consistency of marker homozygosity with the haplotypes across these regions. In addition, mutations within the regulatory regions were not screened for and so cDNA would need to be sequenced for each gene in order to detect this and any alternative transcripts. It should be noted, however, that the eleven databases of ciliated and non-ciliated organisms were not all complete so the apparent absence of genes or proteins does not necessarily mean they are truly not present in that organism.

Of the two leucine rich repeat (LRR) candidate genes identified and sequenced in this chapter (*LRRC35* and *LRRC46*), *LRRC46* was considered the more promising candidate for carrying a potential disease-causing mutation due to greater homology with LRRC50 and better linkage verification. *LRRC50* is the human orthologue of *Chlamydomonas ODA7*, an axonemal OID linker protein, which is defective in the *Chlamydomonas oda7* slow swimming mutant with ultrastructural abnormalities similar to those observed in the presented PCD families. The zebrafish *lrrc50* mutant, *hu255H*, has ciliary defects affecting the dynein arms, also similar to the phenotype observed in the Pakistani families. It is therefore possible that mutations in the *LRRC46* gene could cause disruption of the I/ODA complex, consistent with the Pakistani family ultrastructural phenotype of missing inner and outer dynein arms, and thereby cause PCD. Although no mutations were identified in either gene *LRRC35* or *LRRC46* at the genomic level, it is still possible the defects in transcription or splicing are present and were missed by sequencing only 20 bp into the introns or 40 bp into the 5' and 3' UTRs. However, cDNA was not available in family 141 in order to confirm this.

In summary, two indistinguishable genetic loci were identified by a linkage analysis and homozygosity mapping strategy for PCD with absence of inner and outer dynein arms in five consanguineous families originating from Pakistan. Four good positional candidate genes with putative roles in the structure and/or function of cilia were screened for open reading frame mutations in linked families within the two critical regions and no mutations were detected. Genetic heterogeneity and lack of family material would make it difficult to determine which is the true locus responsible for

the disease in the three families with one affected child. However, the two family pedigrees with three affected offspring are well supported for linkage at these two loci.

4 CHAPTER FOUR

4.1 *LINKAGE ANALYSIS AND MUTATION SCREENING IN TWO ARABIC PCD FAMILIES*

This chapter describes a high density SNP-based genome wide linkage screen, multipoint linkage analysis and homozygosity mapping used to identify the causative gene in an inbred Bedouin PCD family, 152. The consanguineous marriage of family 152 produced four affected (one child died so no DNA sample was obtained) and one unaffected offspring, the latter has one affected child and seven unaffected in another consanguineous marriage, where the husband is the grandson from her grandfathers second wife (Section 2.3.3). The family had classic symptoms of PCD and were excluded for other phenotypes including immunodeficiency and cystic fibrosis (CF) and had abnormal cilia motility. No gross EM defect had been found however, and none of the five affected individuals had *situs inversus*. This chapter describes the identification of a new PCD locus in this family that was found to overlap with a locus consistent for linkage in another Arabic PCD family, 146. Family 146 are a first cousin consanguineous marriage with three affected and four unaffected offspring (Section 2.3.2), from the United Arab Emirates (UAE), with no incidence of *situs inversus* and an ultrastructural defect of intermittent central pair agenesis (Stannard et al., 2004). Analysis of positional candidate genes at the common locus identified on chromosome 6 revealed a shared mutation between both families in an axonemal structural component, *C6ORF206* carried in homozygous state by all affected children.

4.1.1 Power simulation for family 152

For the purposes of power calculations and linkage analysis the complicated structure of family 152 was simplified as shown in Figure 4.1. A high density genome scan was performed in family 152 to determine the disease locus. Prior to this, the power of family 152 to statistically prove linkage was evaluated using linkage simulation programme SLINK (Lemire, 2006), in the same manner and using the same parameters as that described in Chapter 3 Section 3.1.1. The maximum LOD score that could be obtained under a model of homogeneity for this pedigree was calculated. To this end, minimum and maximum achievable LOD scores for family pedigree 152 were analysed. The maximum LOD score that could be achieved

directly at the disease locus ($\theta = 0$) was 3.8 for family 152 (Table 4.1). This analysis indicated that family 152 would provide sufficient linkage power alone.

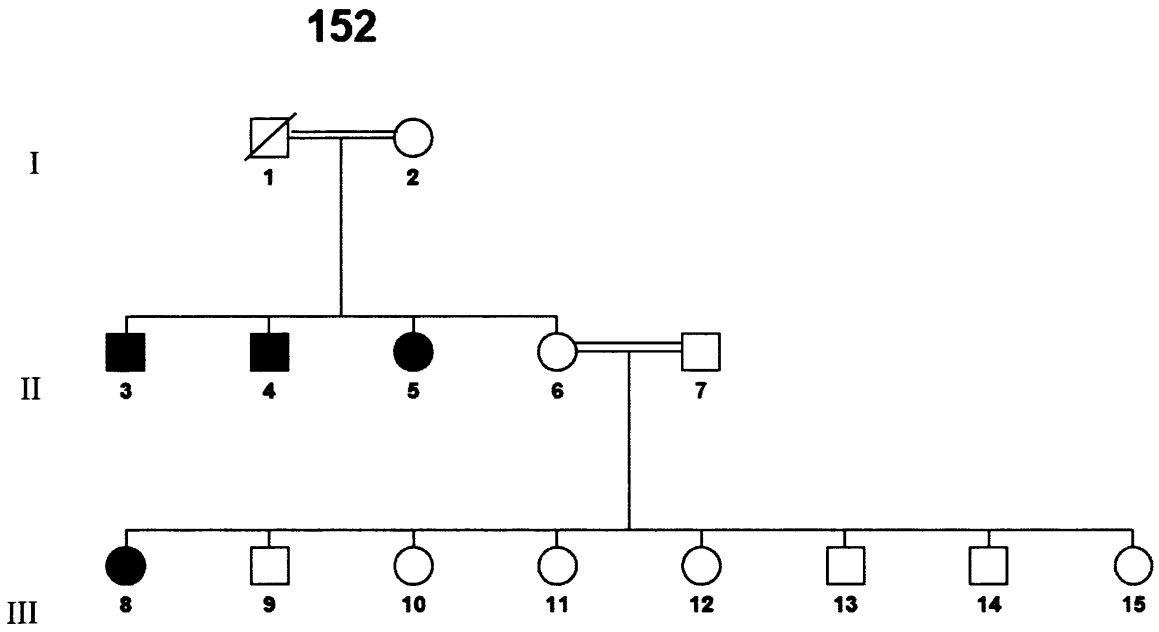


Figure 4.1: Pedigree structure with reduced number of loops for family 152 used in SLINK and MERLIN linkage analysis programmes. Double line indicates consanguinity (first cousin marriages), circles indicate female, squares indicate male. Black shading represents affected individuals, white shading represent non-affected.

	Family 152	
θ	Min	Max
0	-13.52	**3.78**
0.05	-1.7	3.43
0.1	-1.02	3.07

Table 4.1: SLINK simulation of minimum and maximum achievable LOD scores for family 152. Numbers in bold highlight the maximum LOD score achievable at $\theta = 0$ for family 152.

4.1.2 Genome-wide SNP-based linkage screen performed in family 152

A 600 kb density SNP screen (*Illumina* Linkage IVb panel) was performed on DNA samples from family 152, comprising both grandparents (generation I unaffected), all three affected and one unaffected offspring (152 II:6, figure 4.1), the husband of the unaffected offspring (152 II:7) and their affected child (152 III:6). Multipoint parametric linkage analysis was carried out on the *Illumina* SNP genotype data using the MERLIN programme (Abecasis et al., 2002), assuming autosomal recessive inheritance, a disease allele frequency of 0.007 and a disease penetrance of 0.9. Since the MERLIN programme software could not cope with the large complicated family 152 pedigree structure (Section 2.3.3), a simpler model as seen in Figure 4.1 was used to simulate the consanguineous relationships resulting in affected children in this analysis. This revealed only one locus with a significant LOD score >3, on chromosome 6p21.2-21.1 between SNP markers *rs1738240* and *rs945131* (Table 4.2).

Chromosome	Family 152 LODmax (*Illumina* SNPs only)
1	0.551
2	-3.115
3	-1.769
4	-3.881
5	-1.724
6	**3.638**
7	-4.850
8	-3.421
9	-3.033
10	-2.872
11	-0.268
12	-3.531
13	-3.285
14	-2.559
15	-0.189
16	-6.463
17	-0.756
18	-5.010
19	-2.750
20	-4.366
21	-3.264
22	-2.454

Table 4.2: Maximum LOD scores calculated for family 152 using MERLIN linkage analysis on each chromosome. A single region with a significant LOD score of >3 was revealed on chromosome 6 (bold). No other region exceeded 0.5.

4.1.3 Refinement of chromosome 6 locus in Arabic family 152

In order to maximise genetic information and with the aim of refining the chromosome 6 region of interest, 23 microsatellites were genotyped across all individuals in family 152, including the seven non-affected siblings of individual 152 III:8. In order to place the 10 in-house and 13 known microsatellite markers based on their physical positions, a composite genetic map was constructed based on the *Illumina* genetic map and the integrated Nievergelt genetic map, both of which are based on deCODE marker positions (Nievergelt et al., 2004).

4.1.4 Multipoint linkage analysis on Arabic families after additional genotyping

Multipoint linkage analysis was performed on the chromosome 6 region of interest consistent for linkage in family 152 using the more dense genotype data and MERLIN software, assuming the same parameters in section 3.1.2. Limitations of the MERLIN software meant that only eleven people could be used in the analysis, therefore three extra non-affected individuals (152 III:9, III:10 and III:11) were chosen in addition to all parents and affected offspring.

After the additional marker-typing a maximum LOD score of 4.2 was achieved for family 152 between microsatellite marker *rs722269* and SNP marker *rs3734693* (Table 4.3 and Figure 4.2).

Marker name	Genetic distance (cM)	Family 152 - Illumina screen SNPs only LOD score	Family 152 - Illumina screen SNPs plus extra microsatellite markers LOD Score
D6S291	55.51		-0.327
TEX27	56.00		-0.818
D6S389	57.25		-1.227
rs1738240	57.62	-6.150	-5.745
DNAH8 IN 35	57.72		-5.890
D6S1610	58.47		-4.289
rs1885615	58.53	3.369	0.440
D6S2427	58.62		0.753
D6S1562	60.04		1.075
rs715831	60.07	3.627	1.059
D6S1607	60.48		-4.391
D6S1616	60.61		-4.391
D6S426	60.69		-0.034
rs736794	61.31	3.638	-0.051
D6S1575	61.58		-4.390
rs629832	61.75	3.621	0.126
rs1001015	63.29	3.527	0.498
D6S1552	63.96		-4.507
D6S400	64.29		-4.825
rs722269	64.36	3.409	2.793
TBCC	64.86		**3.826**
KNSL8	65.16		**4.130**
rs1537638	65.87	3.089	**4.244**
D6S1582	65.87		**4.244**
UAEPal1	65.97		**4.246**
UAEPal4	66.00		**4.246**
UAUPal2	66.10		**4.246**
rs1563788	66.19	2.970	**4.246**
rs2277123	66.40	2.868	**4.246**
UAEPal5	66.50		**4.247**
UAEPal3	66.60		**4.247**
D6S1604	66.73		**4.247**
rs945131	66.79	2.575	**4.247**
UAEPal6	66.83		**4.247**
D6S451	66.90		**4.247**
rs3734693	67.19	-0.091	-Infinity
D6S2438	68.00		0.482
D6S459	70.14		-Infinity
D6S1638	72.10		-Infinity

Table 4.3: Multipoint LOD scores across chromosome 6p locus in family 152. LOD scores shown for family 152 after *Illumina* SNP screen then further refinement with microsatellite marker-typing. Microsatellites are shown in bold. LOD scores >3 are highlighted in red. Genetic distance in centimorgans (cM) as determined by a composite genetic map constructed based on the *Illumina* genetic map and the integrated *Nievergelt* genetic map, both of which are based on deCODE marker positions. In-house markers were placed according to their physical positions.

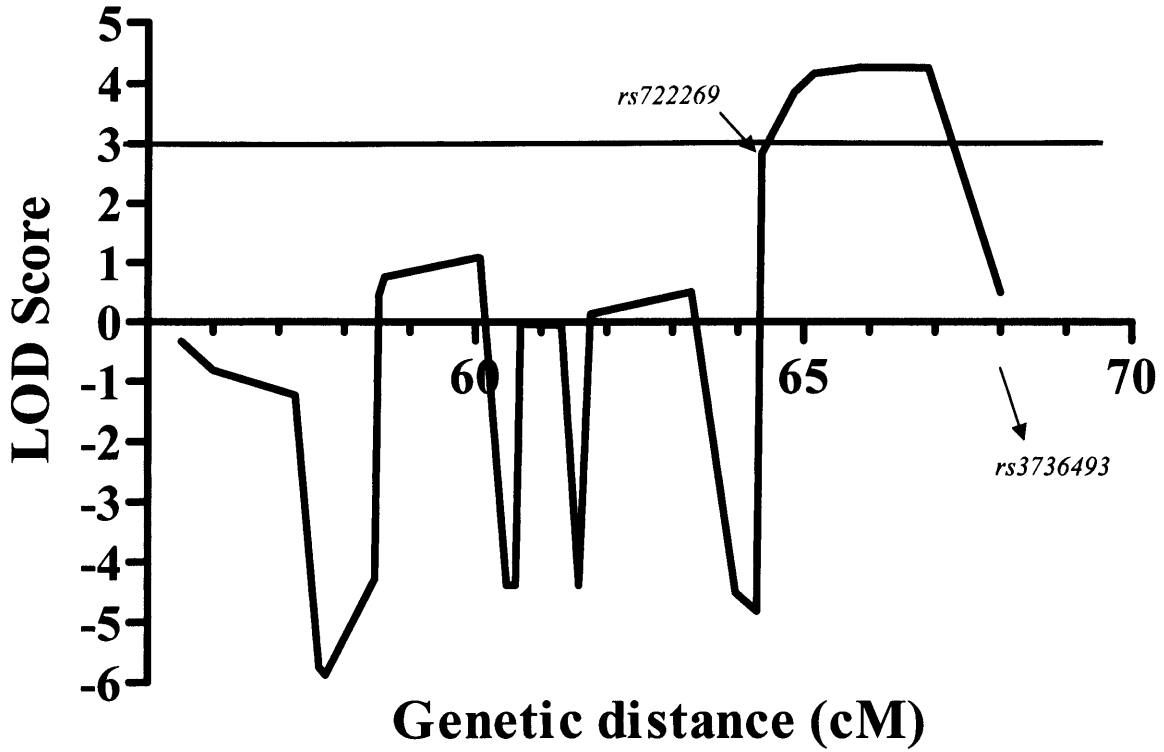


Figure 4.2: Multipoint LOD score across chromosome 6p in family 152. Maximum LOD of 4.2 achieved between markers *rs722269* and *rs3734693* (2.8 cM).

4.1.5 Haplotype analysis of locus identified on chromosome 6p in family 152

To analyse regions of homozygosity, haplotypes were created for the region of interest on chromosome 6p in family 152, using Haplopainter (Thiele and Nurnberg, 2005).

The region consistent with linkage in family 152 had a 2.9 cM (1.8 Mb) region of homozygosity from *D6S400* to *rs3734693*. Haplotypes revealed that this region was defined by loss of homozygosity in affected individuals of family 152 at marker *D6S400* and a recombination event in individual 152 II:3 at SNP marker *rs3734693* (Figure 4.2). Interestingly, genotyping of unaffected siblings of affected individual III:8 in family 152 showed that none of the seven unaffected individuals shared a haplotype with the affected offspring and that the non-affected parent II:6 carried a putative disease chromosome which had been inherited by her affected child, III:8 (Figure 4.3). This data confirmed the chromosome 6 region as a potential disease locus in family 152. In addition, there was some intermittent homozygosity telomeric to this region but this was not considered significant because it did not achieve a significant LOD score in this family (Table 4.3 and Figure 4.3).

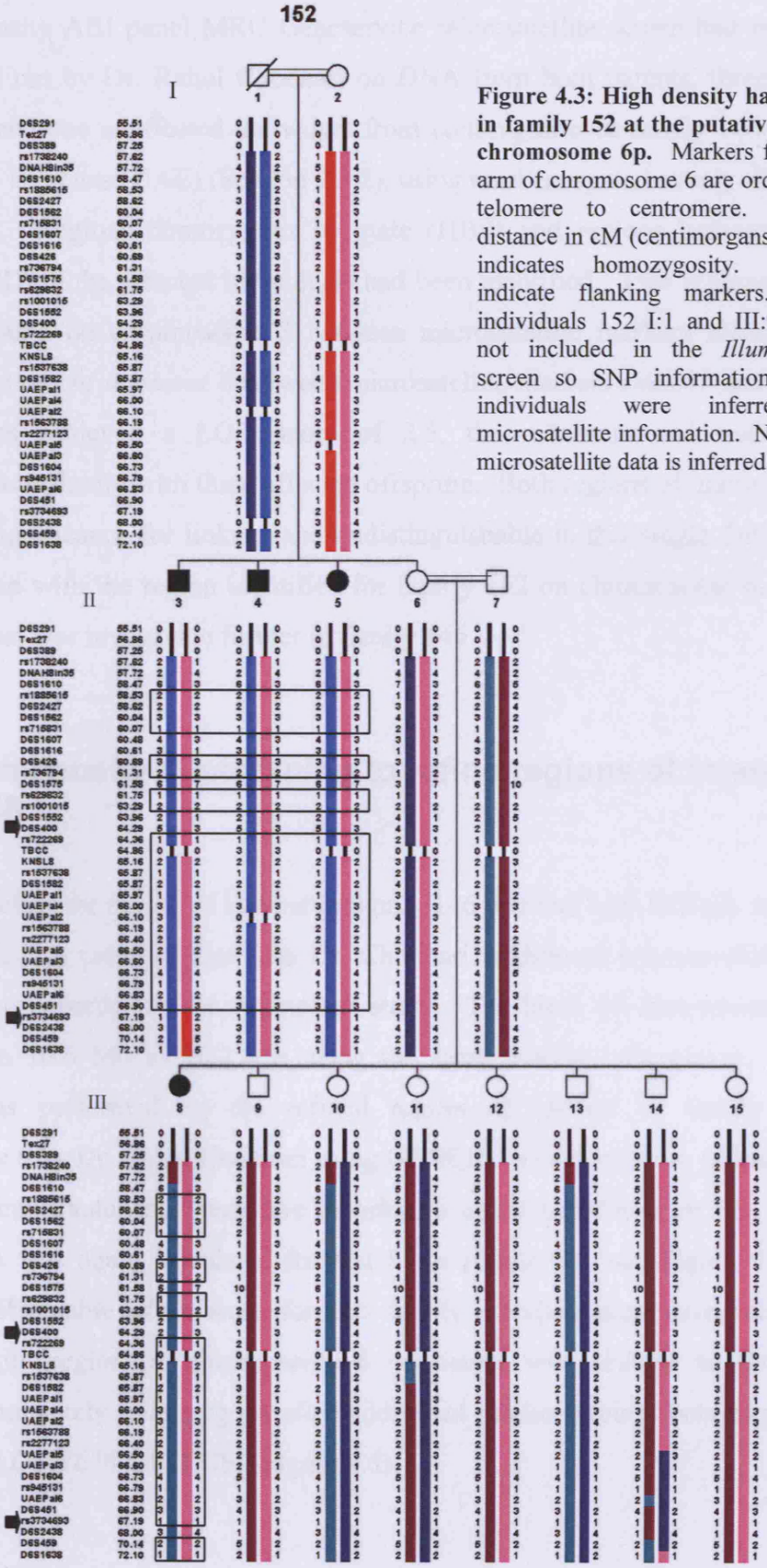


Figure 4.3: High density haplotyping in family 152 at the putative locus on chromosome 6p. Markers from the p arm of chromosome 6 are ordered from telomere to centromere. Genetic distance in cM (centimorgans). Boxing indicates homozygosity. Arrows indicate flanking markers. Note, individuals 152 I:1 and III:9-15 were not included in the *Illumina* SNP screen so SNP information in these individuals were inferred from microsatellite information. None of the microsatellite data is inferred.

4.1.6 Shared homozygous region of interest on chromosome 6 in Arabic family 146

A 10cM density ABI panel MRC Geneservice microsatellite screen had previously been carried out by Dr. Rahul Chodhari on DNA from both parents, three affected individuals and one unaffected individual from consanguineous family 146 from the United Arab Emirates (UAE) (Section 2.3.2), using markers spread evenly throughout the genome. Regions homozygous by state (HBS) and regions homozygous by descent (HBD) in the affected individuals had been identified. Two regions of HBD had been found on chromosome 3 between microsatellite markers *D3S2338* and *D3S1289* and on chromosome 6 between microsatellite markers *D6S291* and *D6S452*. Both regions achieved a LOD score of 2.5, the maximum achievable for a consanguineous family with three affected offspring. Both regions of interest were of equivalent significance for linkage and undistinguishable in this single family. Due to the overlap with the region identified for family 152 on chromosome 6, this was the region that was investigate further in family 146.

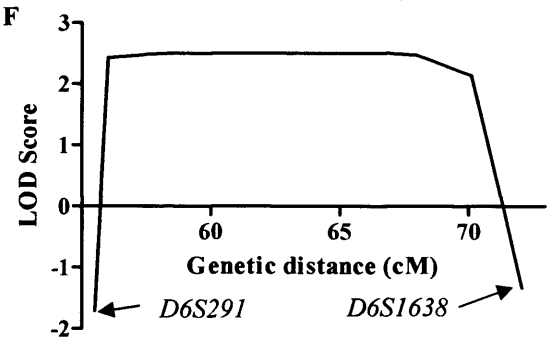
4.1.7 High density haplotyping to refine regions of interest in family 146

In order to refine the region of interest in family 146 that had been initially mapped to chromosome 6 in collaboration with Dr. Chodhari, additional microsatellite marker genotyping was performed in the current study. The locus on chromosome 6 was refined from 10.6 Mb to 10.2 Mb using the extra marker information. Linkage analysis was performed on the refined region of interest in family 146 on chromosome 6 by Dr. Rahul Chodhari using GENEHUNTER software (Strauch et al., 2000), assuming autosomal recessive inheritance and a penetrance of 0.9. A LOD score of 2.5 was again calculated for this locus (Table 4.4 and Figure 4.4A); the maximum obtainable LOD score for this family. Haplotyping revealed that the higher density region on chromosome 6 consistent with linkage to family 146 remained completely homozygous after additional marker typing, between markers *D6S291* and *D6S1638* (10.2 Mb) (Figure 4.5).

		Family 146 - all microsatellite data available
Marker name	**Genetic distance (cM)**	**LOD Score**
D6S291	55.51	-1.710
TEX27	56.00	2.431
D6S389	57.25	2.472
rs1738240	57.62	2.490
DNAH8 IN 35	57.72	2.494
D6S1610	58.47	2.502
rs1885615	58.53	2.502
D6S2427	58.62	2.502
D6S1562	60.04	2.502
rs715831	60.07	2.502
D6S1607	60.48	2.502
D6S1616	60.61	2.502
D6S426	60.69	2.502
rs736794	61.31	2.502
D6S1575	61.58	2.502
rs629832	61.75	2.502
rs1001015	63.29	2.502
D6S1552	63.96	2.502
D6S400	64.29	2.502
rs722269	64.36	2.502
TBCC	64.86	2.502
KNSL8	65.16	2.502
rs1537638	65.87	2.502
D6S1582	65.87	2.502
UAEPal1	65.97	2.502
UAEPal4	66.00	2.502
UAUPal2	66.10	2.502
rs1563788	66.19	2.502
rs2277123	66.40	2.502
UAEPal5	66.50	2.502
UAEPal3	66.60	2.502
D6S1604	66.73	2.502
rs945131	66.79	2.502
UAEPal6	66.83	2.501
D6S451	66.90	2.501
rs3734693	67.19	2.493
D6S2438	68.00	2.471
D6S459	70.14	2.139
D6S1638	72.10	-1.325

Table 4.4: Multipoint LOD scores across chromosome 6p locus in family 146

Figure 4.4: LOD plot for region of interest on



chromosome 6 identified for family 146. Maximum LOD score of 2.5 between microsatellite markers *D6S291* and *D6S1638*.

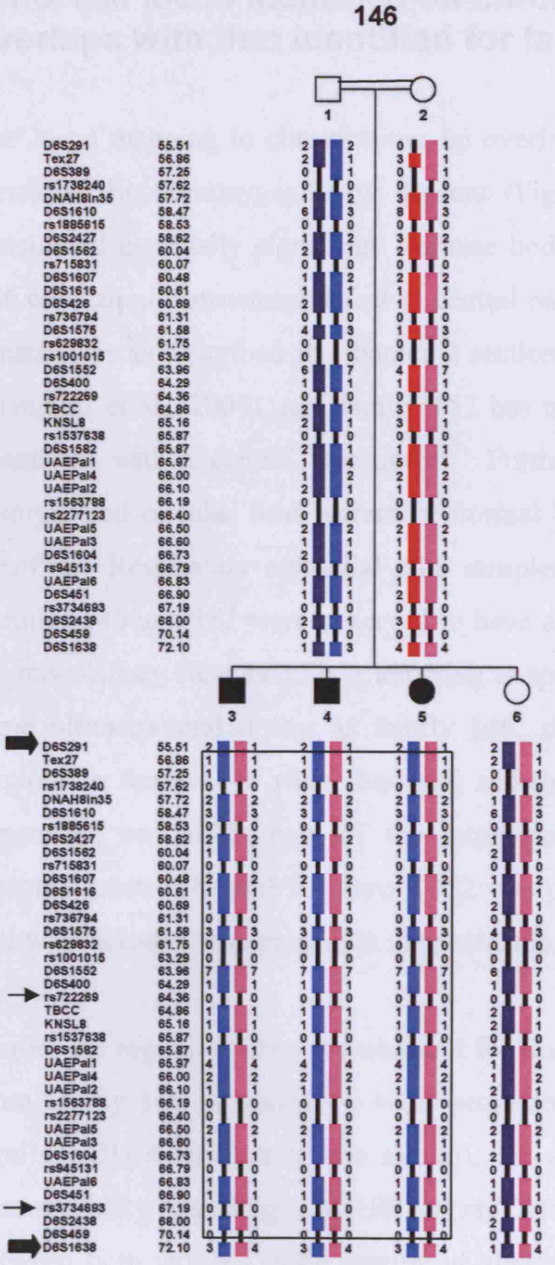


Figure 4.5: Haplotype for region of interest on chromosome 6 identified for family 146. Haplotype shows region of interest defined by loss of homozygosity at microsatellite markers *D6S291* and *D6S1638*. Markers from the p arm of chromosome 6 are shown in order from the telomere to centromere. Genetic distance in cM (centimorgans). Boxing indicates homozygosity. Flanking markers highlighted with bold arrows, flanking markers for family 152 indicated with smaller arrows. *Illumina* SNPs were not typed in this family but included in the haplotype for ease of comparison with family 152 haplotype (Figure 3).

4.1.8 The locus identified on chromosome 6p in family 146 overlaps with that identified for family 152

The locus mapping to chromosome 6p overlapped in both families 146 and 152 as identified by homozygosity by descent (Figure 4.3 and 4.5). This overlap was considered especially significant because both families are of Arabic origin, family 146 have an ultrastructural defect of central pair agenesis with intermittent absence of central pair as described in Chapter 2 section 2.3.2 and no *situs inversus* is present (Stannard et al., 2004), and family 152 has no incidence of *situs inversus* which is consistent with a central pair defect. Furthermore, the cilia motility defect of a disorganised circular beat pattern of normal beat frequency was consistent in both families. Respiratory epithelial cilia samples from affected individuals from both families 146 and 152 were observed to have a circular beat pattern that is ineffective for mucociliary clearance. It is tempting to speculate that family 152 in fact have the same ultrastructural defect as family 146, since only 11.0 - 17 % of cilia cross sections in family 146 were abnormal and this abnormality could easily be missed depending on which part of the axoneme was sectioned for EM in clinical ascertainment. The EM for family 152 was performed in the Israeli clinical setting and was recorded as normal cilia ultrastructure.

Across the region of interest consistent for linkage in family 152, the additional data from family 146 increased the LOD score from 4.2 to 6.7 across the region (Joint family LOD score data is not shown). It was noted that although the region of homozygosity spanning *D6S400* to *rs3734693* on chromosome 6p was shared between both families, little sharing of alleles was evident between family 146 and family 152. Allele sharing is an indication of linkage disequilibrium and a common disease mutation. However, two markers *UAEPal3* and *UAEPal5* did share the same alleles in both families, forming the shared haplotype "2-1" between flanking markers *rs2277123* and *D6S1604*, a region of 212 kb.

4.1.9 Prioritisation of positional candidate genes at the locus identified on chromosome 6 consistent with linkage in families 146 and 152

Candidate genes were identified in the same way as for the Pakistani loci described in Chapter 3 Section 3.1.7. Briefly, a comparative bioinformatics approach was used to identify positional candidate genes in the genetic critical region consistent with linkage identified on chromosome 6, between microsatellite markers *D6S400* and *rs3734693*. This is the region shared by both families.

4.1.9.1 Positional candidate genes at the chromosome 6 locus consistent for linkage in families 146 and 152

Comparative bioinformatic analysis identified the positional candidate genes. See Appendix 2 for full bioinformatic screen. The 1.8 Mb region between *D6S400* and *rs3734693* contained 40 genes in total, 33 with a known function and 7 encoding hypothetical proteins, some of which had a predicted function based on sequence homology. Out of the 33 known genes, 31 were excluded due to their function not being related to cilia or flagella structure or function. The remaining 2 were considered good positional candidates according to their proposed function, bioinformatics BLAST screen, tissue expression and published literature: *TBCC* and *KNSL8*. Out of the 7 hypothetical genes, 6 were initially excluded based on NCBI BLAST homology to non-cilia genes and one was considered a good positional candidate according to results from the comparative BLAST screen (Table 4.5), tissue expression and published literature (Table 4.6): *C6ORF206*.

4.1.9.1.1 *TBCC* (tubulin specific chaperone C)

TBCC (tubulin-specific chaperone C) is located between SNP marker *rs722269* and novel microsatellite *KNSL8* at the telomeric end of the *D6S400 – rs3736493* critical region (Figure 4.6). A novel microsatellite *TBCC* located within the gene had been used in genotyping and the LOD score at this marker was 3.8 (Table 4.3). *TBCC* was considered a good positional candidate as it had homologues conserved to a value within the defined E value of $\leq 10^{-5}$ in all the ciliate datasets in the BLASTp search,

except no homologues were detected in the *Chlamydomonas* Flagellar proteome, the *Chlamydomonas reinhardtii* deflagellation dataset and *Foxj1* ciliogenesis microarray dataset (Table 4.5). It is involved in the process that leads to the correct folding of β-tubulin, since TBCC stimulates the release of tubulin dimers for incorporation into microtubule structures (Kortazar et al., 2007). TBCC has been shown to associate with ARL3, a member of the Arf family of regulatory GTPases that regulate microtubule-dependent processes, and it localises to the connecting cilium in retinal photoreceptor rod cells (Grayson et al., 2002). The photoreceptor connecting cilium structure shares many proteins in common with respiratory cilia (Roepman and Wolfrum, 2007). The NCBI BLASTp search revealed that *TBCC* had homology to the X-linked retinitis pigmentosa gene, *RP2* (Bartolini et al., 2002). Some forms of retinitis pigmentosa are likely to involve IFT defects in the connecting cilium found in rod cells of the retina (Badano et al., 2006). Furthermore, mutations in the *RPGR* gene are associated with PCD and retinitis pigmentosa (Moore et al., 2005). Expression profiling from analysis of EST counts showed widespread human tissue expression for *TBCC*, including expression in eye, kidney, testis and lung (Unigene) (Table 4.6). It was therefore considered that mutations in this gene could cause disruption of ciliary microtubule structures and therefore PCD.

4.1.9.1.2 *KNSL8* (Kinesin-like protein 8)

Kinesin-like protein with microtubular activity, *KNSL8*, is located between novel microsatellite marker *TBCC* and SNP marker *rs1537368* (Figure 4.6). The in-house microsatellite *KNSL8* located within the gene was used in genotyping and the multipoint LOD score at *KNSL8* was 4.1 (Table 4.3). The comparative BLAST screen revealed it was not conserved in any of the non-ciliate datasets but had homology within the defined cut off E value of $\leq 10^{-5}$ to genes in all the ciliate datasets with the exception of the *Chlamydomonas* deflagellation database, *Trypanosoma* flagellar proteome and the *Foxj1* ciliogenesis microarray dataset (Table 4.5). Expression profiling suggested by analysis of EST counts showed *KNSL8* had widespread human tissue expression, including expression in eye, kidney, testis, lung and trachea (Unigene) (Table 4.6). Kinesins are known to function in IFT (Cole, 2005) and therefore this gene could potentially play a role in cilia structure and function.

4.1.9.1.3 *C6ORF206* (chromosome 6 open reading frame 206 hypothetical protein)

C6ORF206 is located between novel microsatellite markers *UAEPal5* and *UAEPal3* where the multipoint LOD score achieved was 4.2 (Figure 4.6) and a putative shared haplotype had been identified in families 146 and 152. This gene had an excellent comparative BLAST profile, with no conservation observed in the three non-ciliate datasets but full conservation in the motile ciliate databases. It was not conserved in *C. elegans*, which only have non-motile cilia (Table 4.5). This suggested that *C6ORF206* could play a specific role in motile cilia. Expression profiling by EST counts revealed that this gene was widely expressed in human tissues, including brain, eye, lung and testis. *C6ORF206* was initially recorded in NCBI and GenBank as a mitochondrial ribosomal-like protein. During this study, homology to an axoneme component *Chlamydomonas reinhardtii* radial spoke protein 9 (*RSP9*) with an E-value of 1e-18, was subsequently revealed by Yang *et al* (Yang et al., 2006) (Table 4.6). RSP9 is a small protein located in the radial spoke head complex of *Chlamydomonas reinhardtii*, adjacent to the central pair (Chapter 1, Section 1.1.4.2.2). In addition, a *Chlamydomonas reinhardtii* RSP9 mutant exists with paralysed flagella (*pf17*), which displays loss of radial spoke heads and central pair displacement (Brokaw and Luck, 1985) (Huang et al., 1981). Therefore *C6ORF206* was an ideal candidate for the central pair agenesis defect seen in EM of cilia cross sections from family 146 (Stannard et al., 2004).

Gene	Chromosome	ARAB	CERV	POMB	CHLY	PAZO	DEFLA	TRYP	CION	TETR	McKE	ELEG	FOXJ1
Tubulin-specific chaperone c (*TBCC*)	6p	0	0	0	1	0	0	1	1	1	1	1	0
Kinesin-like 8 (*KNSL8*)	6p	0	0	0	1	1	0	1	1	1	0	1	0
Chromosome 6 open reading frame 206 (C6orf206)	6p	0	0	0	1	1	1	1	1	1	1	0	1

Table 4.5: Comparative species BLAST results for candidate genes at the chromosome 6 locus consistent for linkage in Arabic PCD families 146 and 152

Key
ARAB - *Arabidopsis thaliana*
CERV - *Saccharomyces cerevisae*
POMB - *Saccharomyces pombe*
CHLY - *Chlamydomonas reinhardtii* genome
PAZO - Pazour's *Chlamydomonas* Flagellar proteome
DEFLA - *Chlamydomonas reinhardtii* deflagellation dataset
TRYP - *Trypanosoma brucei* and *Leishmania major* combined genomes
CION - *Ciona intestinalis* genome
TETR - *Tetrahymena thermophilia* genome
McKE - Gull & McKean's *Trypanosoma* flagellar proteome
ELEG - *Caenorhabditis elegans* proteome
FOXJ1 – Genes down-regulated in *Foxj1* cilia-less mouse

0 = no hit in the BLAST search 1 = hit in the BLAST search

Gene	Chromosome	Function	Domains identified in Human NCBI BLASTp search
Tubulin-specific chaperone c (*TBCC*)	6p	Involved in the pathway leading to correctly folded beta-tubulin from folding intermediates	X-linked retinitis pigmentosa 2 (*RP2*) gene product domain
Kinesin-like 8 (*KNSL8*)	6p	Kinesin motor, motor activity	Kinesin motor domain
Chromosome 6 open reading frame 206 (C6orf206)	6p	Tagged as mitochondrial ribosomal protein S18A-like *	No domains

Table 4.6: Function, domain structure and tissue expression of positional candidate genes on chromosome 6p consistent for linkage in PCD families 146 and 152.). Expression profiling from analysis of EST counts (NCBI). * Homology to RSP9 revealed during the course of this study by Yang et al.

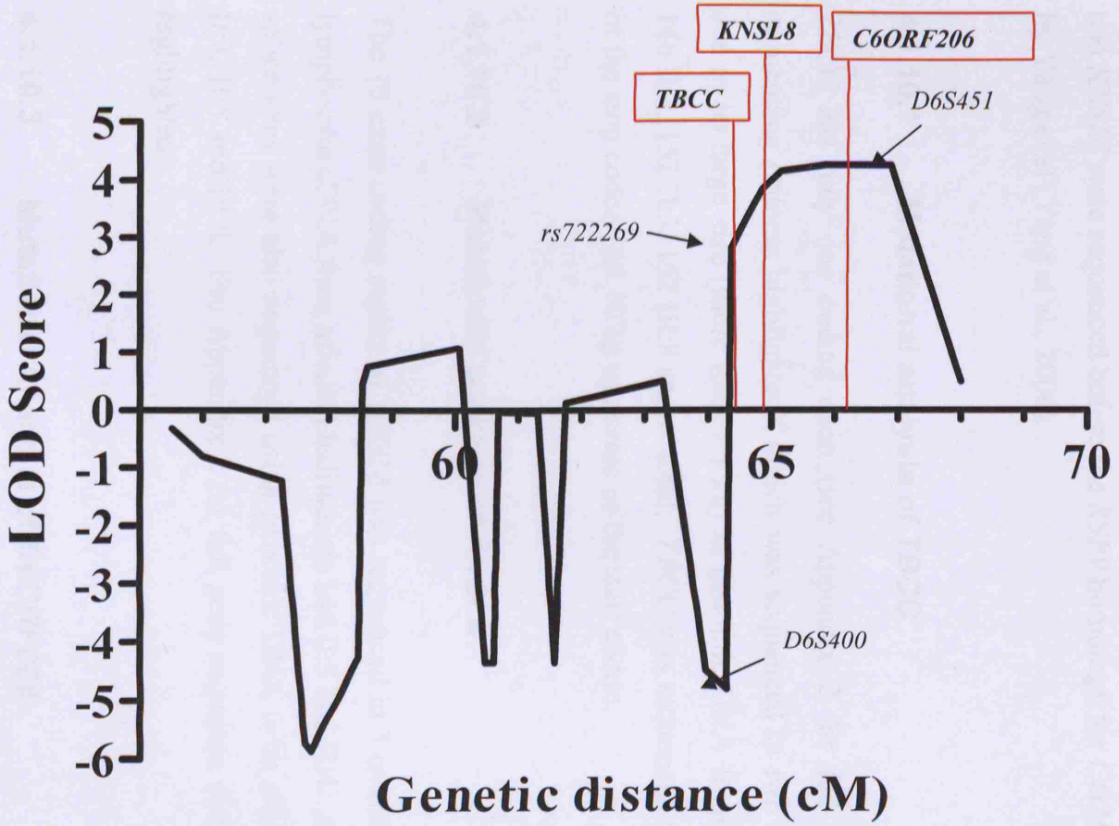


Figure 4.6: LOD plot for chromosome 6p locus consistent for linkage in family 152 with candidate gene locations annotated.

4.1.10 Sequencing of Positional Candidates

Three genes identified as good positional candidates within the locus on chromosome 6 were screened for mutations in affected individuals II:3, II:4 or II:5 of family 146 and individuals II:3 or III:8 of family 152, using primers designed for all coding open reading frames (ORFs), spanning intron-exon boundaries by 20 bp and 40 bp into the 5' and 3' un-translated regions (UTR). Primer sequences are in Appendix 2. *TBCC* and *KNSL8* were sequenced before the *RSP9* homology for *C6ORF206* was published by Yang et al (Yang et al., 2006).

4.1.10.1 Mutational analysis of *TBCC*

TBCC has only one coding exon (see Appendix 2 for full gene sequence with sequencing primers highlighted), which was sequenced in two overlapping sections due to its large size (more than 1.5 kb) in genomic DNA from affected individuals 146 II:3, 152 II:3, 152 III:8 and control. *TBCC* was sequenced to 40 bp downstream of the stop codon and 40 bp upstream of the start codon.

4.1.10.2 Mutational analysis of *KNSL8*

The 15 exon coding region of *KNSL8* was sequenced in 3 overlapping sections using lymphocyte cDNA from affected individuals 146 II:3 and II:4. All 15 exons and their splice sites were also sequenced using genomic DNA from affected individuals 152 II:3, II:5 and III:8. See Appendix 2 for full gene sequence with sequencing primers highlighted.

4.1.10.3 Mutational Analysis of *C6ORF206*

C6ORF206 was sequenced using genomic DNA from affected individuals 146 II:3, 152 II:3, 152 III:8 and control. The full length *C6ORF206* transcript was sequenced in lymphocyte cDNA made from 146 II:4, 146 II:5 and control. A homozygous 3 base pair (bp) deletion in exon 5, 22 bases 5' to the stop codon was identified

(c.804_806delGAA) in all five affected individuals, that was not present in control DNA (Figure 4.7).

This mutation was not present in 126 Bedouin population control matched chromosomes, thus by definition was confirmed to not be a polymorphism.

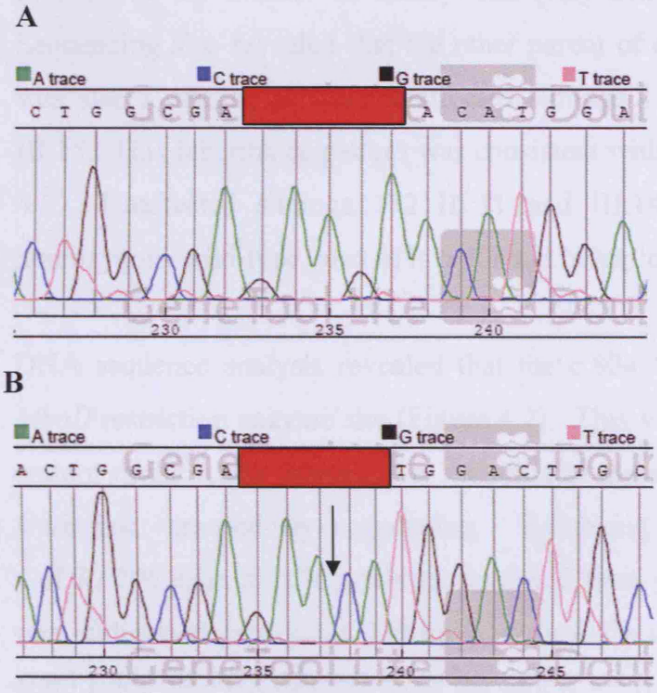


Figure 4.7: Electropherogram of partial sequence of *C6ORF206* exon 5 showing c.804_806delGAA mutation. A) Electropherogram from control individual. B) Electropherogram from individual 152 III:8, black arrow indicates GAA deletion. *MboII* restriction enzyme site destroyed by deletion – highlighted with red box.

4.1.11 Confirming inheritance pattern of *C6ORF206* c.804_806delGAA mutation in PCD families 146 and 152

The c.804_806delGAA mutation identified in *C6ORF206* in DNA from affected individuals of families 146 and 152, was confirmed by sequencing in all affected individuals of both Arabic families: 146 II:3, II:4, II:5 and 152 II:3, II:4, II:5 and III:8. Sequencing of exon 5 of *C6ORF206* showed heterozygosity for the 3 bp deletion in the mother of family 152 (I:2) and her unaffected child (152 II:6). Sequencing also revealed that the other parent of affected individual III:8 (152 II:7) was also a carrier, as were unaffected siblings 152 III:9, III:10, III:12, III:13 and III:15. This inheritance pattern was consistent with the haplotypes in Figures 4.3 and 4.5. Unaffected siblings 152 III:11 and III:14 were also sequenced and was homozygous wild-type, also as is predicted in haplotype analysis (Figure 4.3).

DNA sequence analysis revealed that the c.804_806delGAA mutation destroys an *MboII* restriction enzyme site (Figure 4.7). This was used to confirm the inheritance pattern of the c.804_806delGAA mutation in members of families 146 and 152 who were not screened by sequencing. Following restriction enzyme digestion of *C6ORF206* exon 5 PCR products amplified from genomic DNA from patients, only one undigested product of 350 bp is seen, if the deletion is present in homozygous form since the PCR product lacks the *MboII* restriction enzyme site. If the mutation is not present, the PCR product is digested by *MboII* into two fragments of 259 and 94 bp. All unaffected individuals in both families presented either with two products of 259 bp and 94 bp, demonstrating homozygosity for the wild type allele, or three bands of 350 bp, 259 bp and 94 bp as a result of heterozygosity (Figure 4.8). All parents were heterozygotes for the deletion (Figure 4.8). Just two unaffected individuals in both families were homozygous wild type and did not carry the mutation (152 III:11 and III:14). Thus, restriction enzyme digestion with *MboII* of a PCR fragment of exon 5 of *C6ORF206* in the two Arabic families 146 and 152 supported the haplotype and sequencing results and clearly demonstrated that the c.804_806delGAA mutation perfectly segregated with the disease in both families (Figure 4.8).

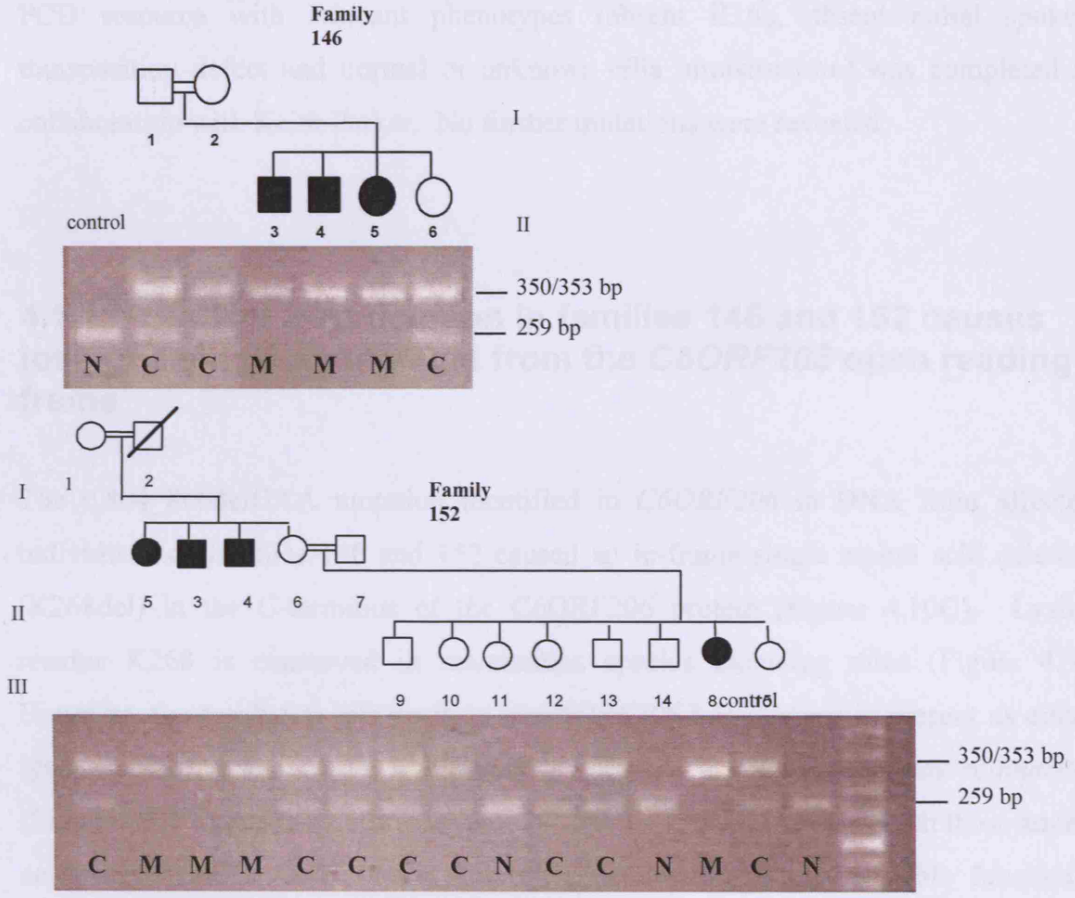


Figure 4.8: ***MboII*** **restriction enzyme digest showing inheritance of the c.804_806delGAA mutation in** ***C6ORF206*** **in families 146 and 152.** N = normal (wild-type homozygous), C = carrier of c.804_806delGAA mutation, M = homozygous mutation.

4.1.12 Screening for further mutations in *C6ORF206* in the PCD family resource

Screening for mutations in all five exons of *C6ORF206* in 39 patients from the UCL PCD resource with relevant phenotypes (absent IDAs, absent radial spokes, transposition defect and normal or unknown cilia ultrastructure) was completed in collaboration with Keith Parker. No further mutations were revealed.

4.1.13 The 3 bp deletion in families 146 and 152 causes loss of a single amino acid from the *C6ORF206* open reading frame

The c.804_806delGAA mutation identified in *C6ORF206* in DNA from affected individuals of families 146 and 152 caused an in-frame single amino acid deletion (K268del) in the C-terminus of the C6ORF206 protein (Figure 4.10C). Lysine residue K268 is conserved in mammalian species including mice (Figure 4.9). However, the residue at this position in *C6ORF206* homologues is present as either lysine in mouse and *Ciona intestinalis* or arginine in *Chlamydomonas reinhardtii* (Figure 4.9). Therefore it is not extremely well conserved, although both these amino acids have similar basic, polar properties, which suggests a possible functional conservation of this position between species rather than absolute amino acid conservation. However, in zebrafish this residue is a methionine (Figure 4.9), which is different to lysine and arginine in that it is neutral and non-polar. Sequencing of the zebrafish *C6ORF206* homologue was performed in genomic DNA isolated from 5 days post fertilization (dpf) zebrafish embryos confirmed that the database deposited sequence was the correct sequence.

In collaboration with Dr. Richard Emes (Centre for Applied Entomology and Parasitology, Keele University) protein modelling was performed to identify changes in the protein that could result from the K268del mutation. The C6ORF206 open reading frame is small without any obvious functional domains, so there are few clues to its function. No obvious changes in protein properties were predicted to occur as a result of loss of this amino acid (Figure 4.11): it is not contained in an obvious structural entity or within a known protein biding motif that may be disrupted by the

loss of the amino acid. As it is a conserved amino acid, it is possible that K268 contributes to the overall structure of this C-terminal region of the protein product and it is the absence of K268 rather than its specific properties that are the key to the mutational effect.

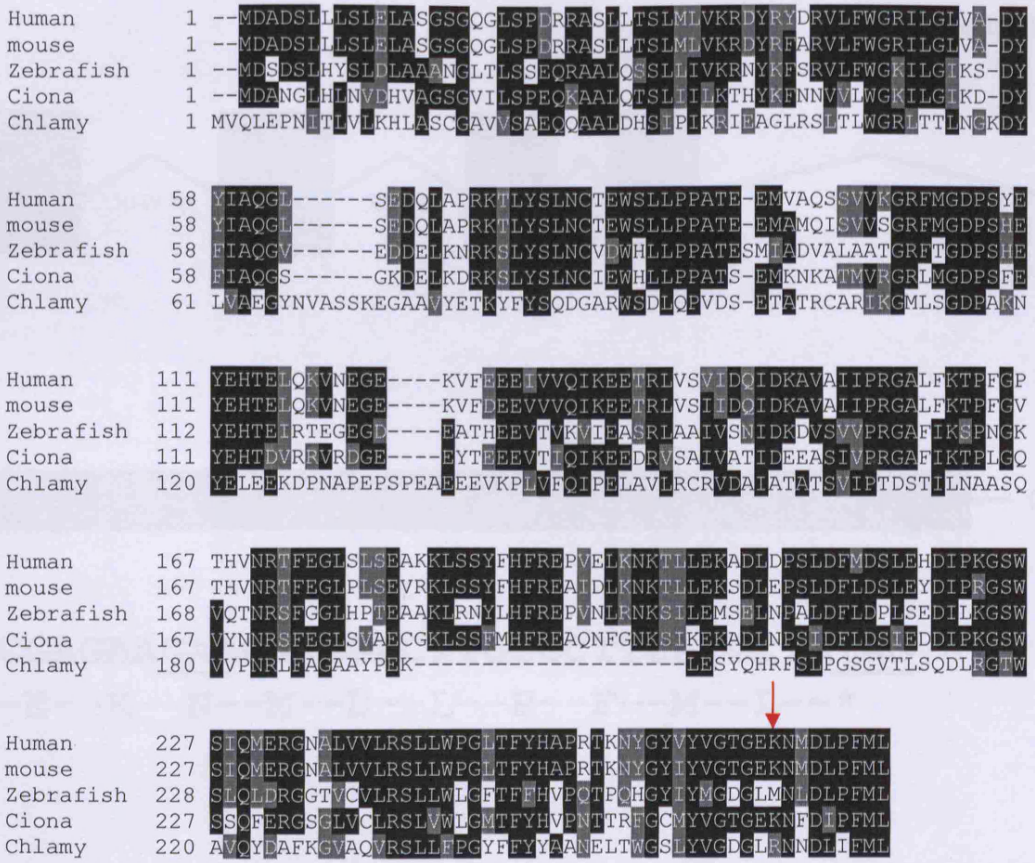
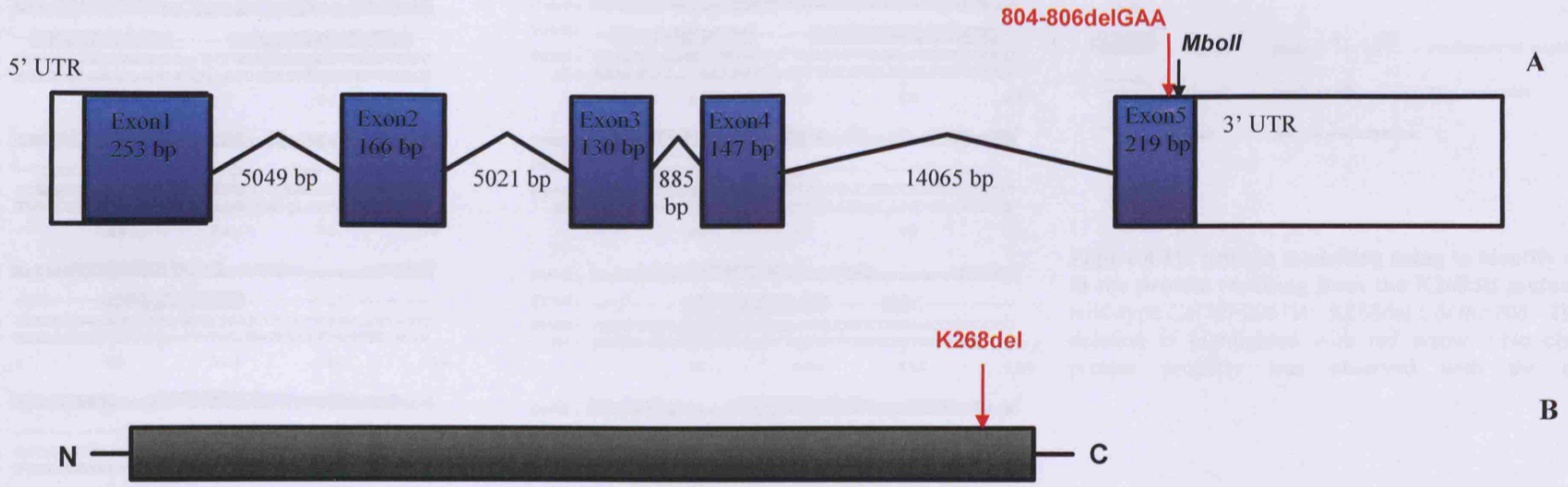


Figure 4.9: Alignment of human *C6ORF206* with mouse, zebrafish, *Ciona Intestinalis* and *Chlamydomonas reinhardtii* homologues demonstrates conservation of residue 268. The red arrow highlights residue 268, which is conserved as either basic, polar residue Lysine (K) or Arginine (R) in *Chlamydomonas,* or Methionine (M) in zebrafish. Accession numbers – Human NP_689945, Mouse NP_083614, zebrafish NP_001025284, *Ciona intestinalis* product of gene JGI 1.0 ID 136430 and *Chlamydomonas reinhardtii* GI:83284713.



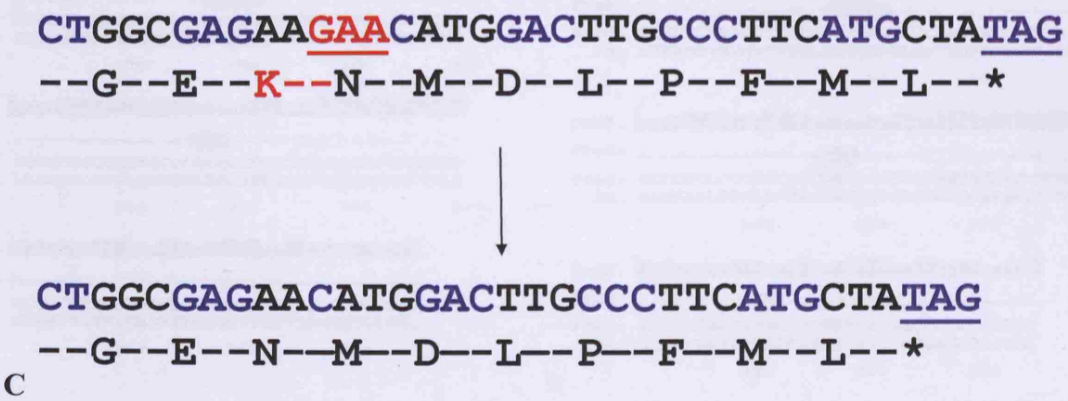


Figure 4.10: *C6ORF206* mutation c.804-806delGAA

A – gene structure of *C6ORF206*, blue boxes = exons, white boxes = UTRs. red arrow indicates location of 3 bp deletion in exon 5 found in PCD patients from Arabic families 146 and 152, black arrow shows *MboII* restriction enzyme site. **B** – C6ORF206 protein, red arrow indicates location of lysine deletion toward the C-terminus of the protein. This protein has no known domains (Yang et al., 2006). **C** - effect of the in-frame c.804_806delGAA deletion on the amino acid sequence of the C6ORF206 protein. Residue K268 is deleted in the mutant protein.

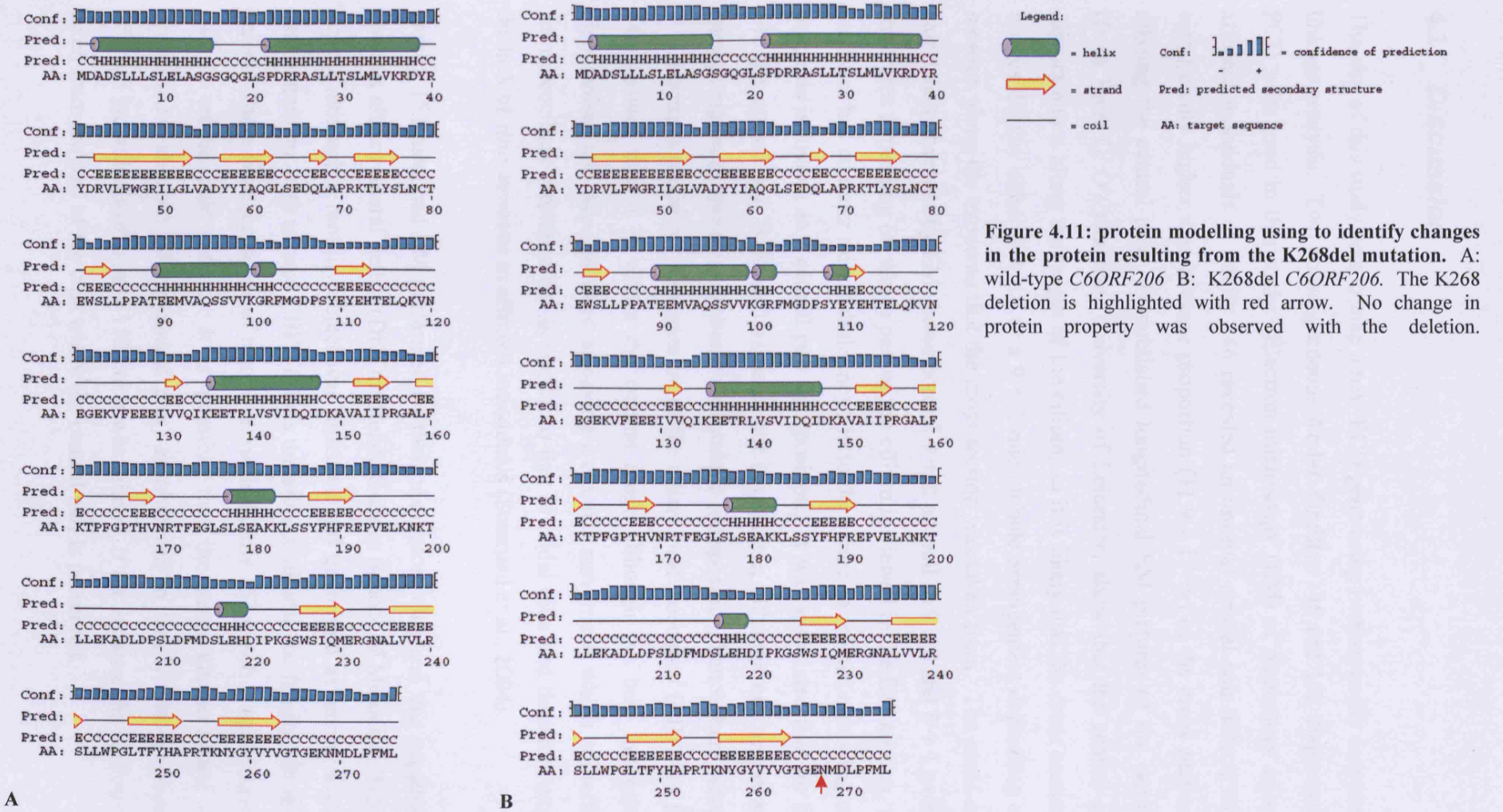


Figure 4.11: protein modelling using to identify changes in the protein resulting from the K268del mutation. A: wild-type *C6ORF206* B: K268del *C6ORF206.* The K268 deletion is highlighted with red arrow. No change in protein property was observed with the deletion.

4.2 *Discussion*

The aim of this study was to map a new PCD gene using homozygosity mapping and linkage analysis. Two consanguineous Arabic families 146 and 152 diagnosed with PCD were used in this study. Electron microscopy (EM) of respiratory cilia from affected individuals of family 146 revealed an unusual central pair defect, where a significantly higher than average proportion (11.9 - 17 %) of the cross sections are missing the central pair. Unpublished longitudinal EM pictures of the same cilia (from Prof. C. O'Callaghan, University of Leicester) show that the central pair is discontinuous along the length of the cilium, so it is likely that the cross section EM images display either a 9 + 0 or a 9 + 2 microtubule arrangement depending on the location along the axoneme that the cross section was taken from. The usual central pair defect observed by EM shows normal 9 + 2, as well as 9 + 0 and 8 + 1 axonemal structures depending on which part of the cilium is viewed in the EM section, where the latter has an outer microtubule move in to the centre to replace a missing central pair, this is known as a central pair transposition. It was speculated that for family 146 the central pair is formed as usual but is unstable, although not across a large enough region to cause a peripheral microtubule transposition, hence there were no 8 + 1 structures observed. Observation of the ciliary beat frequency (CBF) for family 146 showed that it is within the normal range although the beat movement is disorganised and was described as having a circular movement, which is ineffective for mucociliary clearance but is similar to that of nodal cilia and therefore explains the lack of *situs inversus* in affected individuals (Stannard et al., 2004).

Family 152 have had EM performed on their respiratory cilia and this suggested no obvious ultrastructural defect (Dr. Orit Reish, Sackler School of Medicine, Tel-Aviv, Israel), although a motility defect was observed and quantified as having a circular beat pattern but with normal CBF. This is therefore similar to the family 146 motility defect. There is no incidence of *situs inversus* in family 152 which is again consistent with a central pair defect. It is possible that the same ultrastructural defect (intermittent absence of the central pair) as observed in family 146 is present in affected individuals of family 152 but it was missed, if for instance the sections were taken from the part of the cilia where the central pair is preserved.

MERLIN linkage analysis on data from an *Illumina* SNP screen performed on DNA from family 152 identified a locus with a significant multipoint LOD score (Zmax = 3.6) on chromosome 6 between markers *rs1738240* and *rs945131* (Table 4.3, column 2). Microsatellites were typed across the region on chromosome 6 including non-affected siblings. This refined the region of interest from 9.2 cM to 2.9 cM, between flanking markers *D6S400* to *rs3734693*, because the initial region of homozygosity was broken up by heterozygous markers at the telomeric end of the region. Addition of the extra microsatellite data also increased the maximum LOD score from 3.0 to 4.2 between microsatellite marker *rs722269* and SNP marker *rs3734693*, across a region spanning 1.8 Mb.

Analysis of a genome-wide microsatellite screen that had been performed prior to this study had identified two putative disease loci in family 146 of equal weight on chromosome 3 and 6. At both these loci the maximum LOD score was 2.5 since family 146 only has three affected children. The loci separately identified on chromosome 6 in families 146 and 152 were found to overlap between markers *D6S400* and *rs3734693*, a region of 1.8 Mb (2.9 cM). This overlap was considered especially significant because both families are of Arabic origin and as described above had a consistent ultrastructural phenotype.

Combined, the genotype data for Arabic families 146 and 152 yielded a LOD score of 6.7 across the overlapping region of homozygosity on chromosome 6, between markers *D6S400* and *rs3734693*. However, only markers *UAEPal3* and *UAEPal5* had the same alleles across all affected individuals of families 146 and 152, forming the haplotype "2-1" in both families between flanking markers *rs2277123* and *D6S1604*, a region of just 212 kb that contains the candidate gene *C6ORF206*. So although consistent for linkage at the same region, there was no great evidence for linkage disequilibrium (LD) at the available marker density.

Identification of positional candidate genes in the shared critical region on chromosome 6 between markers *D6S400* and *rs3734693* based on conservation in ciliated organisms and any putative cilia structure and/or function and related homology, as well as tissue expression led to prioritization of 3 out of the 40 positional candidate genes present: *TBCC*, *KNSL8* and *C6ORF206*. Sequencing and

restriction digests of *C6ORF206* revealed a homozygous 3 base pair (bp) deletion in exon 5, 22 bases 5' to the stop codon (c.804_806delGAA), in all affected individuals in both families 146 and 152. All parents and non-affected siblings were heterozygous for the mutation, except for two non-affected siblings (152 III:11 and III:14) that were homozygous for the normal allele. Observation of the full family 152 pedigree (Section 2.3.3) revealed that the disease gene is inherited by the three brothers of the first consanguineous marriage (the grandparents of generation II) and therefore is homozygous by descent in affected individuals of generation II. The sharing of the c.804_806delGAA mutation in both families 146 and 152 is interesting as the gene is located between the two markers *UAEPal3* and *UAEPal5* that suggest some allele sharing and therefore linkage disequilibrium. This small shared "2-1" haplotype between flanking markers *rs2277123* and *D6S1604* described in the previous paragraph could have arisen due to uninformative or incomplete parental haplotypes, but subsequent mutational analysis of *C6ORF206* suggested the putative small shared haplotype across the *C6ORF206* gene was real. Alternatively, it could be a mutation hotspot and they share the same mutation by chance, however, this was considered unlikely for a triplet deletion.

The c.804_806delGAA mutation was predicted to cause an in-frame single conserved amino acid deletion (K268del) in the C-terminus of the C6ORF206 protein but no significant alteration of the properties of the protein. However, this is difficult to assess using the available protein structure programs since there are no obvious functional domains in the protein. Further experimental work would be required to analyse the effect of the mutation on protein binding or any other protein properties. For example, expression of the mutation in a fluorescent tagged protein in ciliated cells to observe any changes in localization or function could be performed, or binding properties of the protein examined using an antibody pull down method. It is likely that the deletion of this single amino acid near to the C-terminus of the protein disrupts an important sequence domain that is as yet unidentified, since functional work performed in Chapter 5 showed that this residue is critical for cilia movement.

The *C6ORF206* gene product is homologous to *Chlamydomonas* radial spoke protein 9, which is located in the radial spoke head near the central pair of the axoneme. These two structures have close proximity in the axoneme and mutations in the

C6ORF206 gene are associated with disruption of the central pair. It is also possible that there are additional ultrastructural problems in the radial spokes in these families that are undetectable by EM.

In summary, a gene locus on chromosome 6 was identified by linkage analysis and homozygosity mapping in two consanguineous Arabic families with PCD. Three good candidate genes with putative roles in the structure and/or function of cilia were highlighted within the critical region, one of which was found to contain a mutation that is likely to cause the disease phenotype. The mutation is inherited in homozygous form by all seven affected individuals in families 146 and 152 and segregates perfectly with the disease haplotype in both families. It causes an in-frame deletion of a single amino acid, which is not fully conserved suggesting that it is its loss rather than its chemical properties that cause cilia dysfunction and PCD.

5 CHAPTER FIVE

5.1 *FUNCTIONAL ANALYSIS OF RADIAL SPOKE PROTEIN 9 (C6ORF206/RSP9)*

Following identification of a locus on chromosome 6p21.2-21.1 consistent with linkage in families, 146 and 152, sequencing of positional candidate genes revealed a homozygous 3 bp deletion (c.804_806delGAA) in exon 5 of *C6ORF206/RSP9* in all affected individuals. The c.804_806delGAA mutation was not present in 126 population matched control chromosomes, thus confirming that it was not a polymorphism. The deletion causes an in-frame deletion of a single amino acid (K268del) in the C-terminus of the C6ORF206/RSP9 protein. The affected patients have a rare ultrastructural defect of intermittent absence of the ciliary central pair microtubules (Section 2.3.2). This is also an unusual form of PCD since none of the affected individuals have laterality problems and their respiratory cilia retain a normal beat frequency but with a circular beat pattern similar which is ineffective for mucociliary function. This has been reported as similar to the unidirectional movement pattern of embryonic nodal cilia (Stannard et al., 2004).

This chapter describes functional analysis to investigate the *C6ORF206/RSP9* gene and the functional consequences of the K268del mutation. In order to investigate the tissue expression of *C6ORF206/RSP9*, probes were made for whole mount and section *in situ* hybridisation, which was performed on mouse embryos at stages E7.5 and E19.

In an effort to characterise the function of C6ORF206/RSP9 in a vertebrate model system, the phenotype of zebrafish (*Danio rerio*) embryos injected with gene silencing morpholinos targeting the zebrafish *C6ORF206/RSP9* homologue, was investigated compared to controls.

The *Chlamydomonas* model provides a system to explore the precise functional effect of the c.804_806delGAA mutation. The *Chlamydomonas reinhardtii C6ORF206/RSP9* homologue, radial spoke protein 9 (*RSP9*), is known to be defective in *pf17*, a radial spoke head-deficient *Chlamydomonas* mutant strain which has paralysed flagella. The *C6ORF206/RSP9* gene mutation in *pf17* was determined and studies were conducted to explore the pathological effects of mutations in flagella

motility, comparing the effects of a null allele against that of the human c.804_806delGAA mutation.

5.1.1 Expression pattern of *C6ORF206/RSP9* in the mouse embryo

The expression pattern of the *C6ORF206/RSP9* mouse homologue mRNA in wild-type embryos at different stages of development was analysed, using *DNAH5* as a control probe expressed in the axoneme of motile cilia. The *C6ORF206/RSP9 in situ* probe was made as described in Section 2.4.14 and the *DNAH5* probe was kindly provided by Dr Heymut Omran (Olbrich et al., 2002). The *in situ* work was performed in collaboration with Dr. Nick Greene, UCL. At E19, significant expression of *C6ORF206/RSP9* and *DNAH5* was seen in body sagittal sections in the tracheal epithelium (Figure 5.1A-D) and head sagittal sections revealed *C6ORF206/RSP9* and *DNAH5* expression in the nasal epithelium (Figure 5.1E-H). This provided evidence that *C6ORF206/RSP9* mRNA is expressed in ciliated tissues and this is consistent with a role in cilia motility. In trachea and nasal epithelium the expression overlapped with the known PCD gene *DNAH5*. Expression was also observed in the E7.5 stage embryonic node (Figure 5.1I), suggesting that the gene product is present in nodal cilia. This again is the same pattern seen for the PCD gene *DNAH5* (Olbrich et al., 2002). Since there is no incidence of *situs inversus* in either PCD family 146 or 152, who have the *C6ORF206/RSP9* mutation identified in Chapter 4, it is possible that this gene, unlike *DNAH5*, has a redundant role at the embryonic node.

The presence and level of expression of *C6ORF206/RSP9* mRNA was also monitored using RT-PCR on cDNA made from multiple adult mouse tissues, including lung, trachea, kidney, brain, testis and liver, and actin primers as a housekeeping gene expression control. Strong expression was observed in testis and lower, but marked expression levels were seen in all other tissues in the analysis (data not shown).

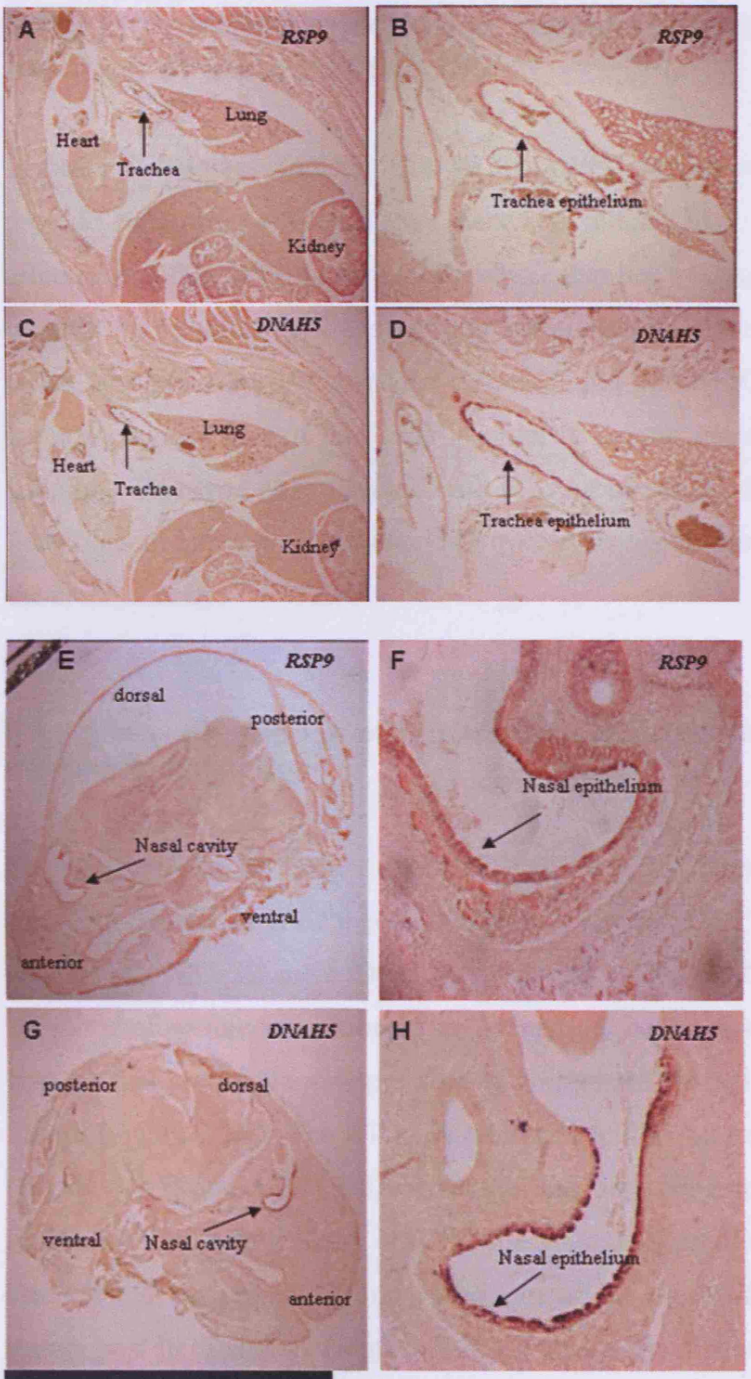


Figure 5.1: *C6ORF206/RSP9* and *DNAH5* expression in whole mount and section *in situ* hybridisation of mouse embryos. A-D: *RSP9* and control *DNAH5* expressed in trachea epithelium in E19 mouse (sagittal sections). B and D close ups of A and C, respectively. **E-H:** *RSP9* and control *DNAH5* expressed in nasal epithelium in E19 mouse head (sagittal sections). F and H close ups of E and G, respectively. **I:** *RSP9* expressed in E7.5 mouse embryonic node (white arrow).

5.1.2 Morpholino injection in zebrafish larvae to assess function of *C6ORF206/RSP9*

In order to address whether *C6ORF206/RSP9* plays a role in cilia function, and provide supporting evidence that the c.804_806delGAA mutation identified in all affected individuals of the two Arabic PCD families in chapter four could be disease-causing, morpholino antisense oligonucleotides (MO) were designed that targeted splice acceptor sites of the second (MOex2) and third (MOex3) coding exons of the zebrafish homologue of this gene in order to disrupt protein function by gene mis-splicing. Embryos were injected with 6.25 ng of either MO at the one-cell stage of development in conjunction with GFP mRNA to ensure proper uptake of the MO into the embryo body.

5.1.2.1 Disruption of zebrafish *C6ORF206/RSP9* mRNA splicing by morpholino injection

The location of both exon 2 (MOex2) and exon 3 (MOex3) MOs in the five exon zebrafish *C6ORF206/RSP9* gene is shown in figure 5.2A. The effectiveness of each morpholino, MOex2 and MOex3, was verified by RT-PCR using RNA isolated from 10 morpholino-injected embryos at 24 and 72 hours post-fertilisation (hpf). This showed that morpholino suppression of correct mRNA splicing persisted for at least 72 hpf (Figure 5.2E). The RT-PCR generated an 837 bp fragment in control embryos (Figure 5.2 B and 5.2E) and correct splicing was confirmed by sequencing of gel-purified RT-PCR fragments. RT-PCR of MOex2-injected embryos analysed with the same primer set produced three amplicons of 1.2 kb, 1.0 kb and 773 bp at all times tested post fertilisation (Figure 5.2E), caused by disrupted splicing of intron 2. Sequencing of the RT-PCR products revealed that this resulted from non-splicing (1.2 kb product) and partial splicing of intron 2 (1.0 kb product), or deletion of 64 bp from the 3' end of exon 2 (773 bp product) presumably due to use of a cryptic splice donor in exon 2 (Figure 5.2C). The predicted results of mis-splicing was premature truncation due to a frameshift in all cases. For the 1.2 and 1.0 kb products, it was expected that only 155 amino acids are expressed and for the 773 bp product, 111 amino acids are expressed. The wild-type zebrafish C6ORF206/RSP9 protein is 277 amino acids in length, thus the truncated proteins were expected to be non-functional. No wild-type mRNA was expressed after MOex2 injection.

RNA from MOex3-injected embryos were analysed in the same way by RT-PCR, which showed three amplicons of 910 bp, 805 bp and 702 bp at all times tested post fertilisation (Figure 5.2E), caused by disrupted splicing of intron 3. Sequencing of the RT-PCR products, revealed partial splicing of intron 3 (910 bp product) where a cryptic splice site is presumably exists, a 28 bp deletion at the 3' end of exon 3 (805 bp product) and complete deletion of exon 3 (702 bp product). The predicted results of the mis-splicing observed in the 910 and 805 bp product is premature truncation due to a frameshift, resulting in predicted 205 and 187 amino acid products. The predicted product of the 702 bp transcript is a full length protein with a frameshift after amino acid 132 of the 277 amino acid protein. No wild-type *C6ORF206/RSP9* mRNA was detected in embryos injected with MOex3. Thus, injection of 6.25 ng per embryo of either splice-blocking morpholino, MOex2 or MOex3, leads to a substantial decrease in functional gene product.

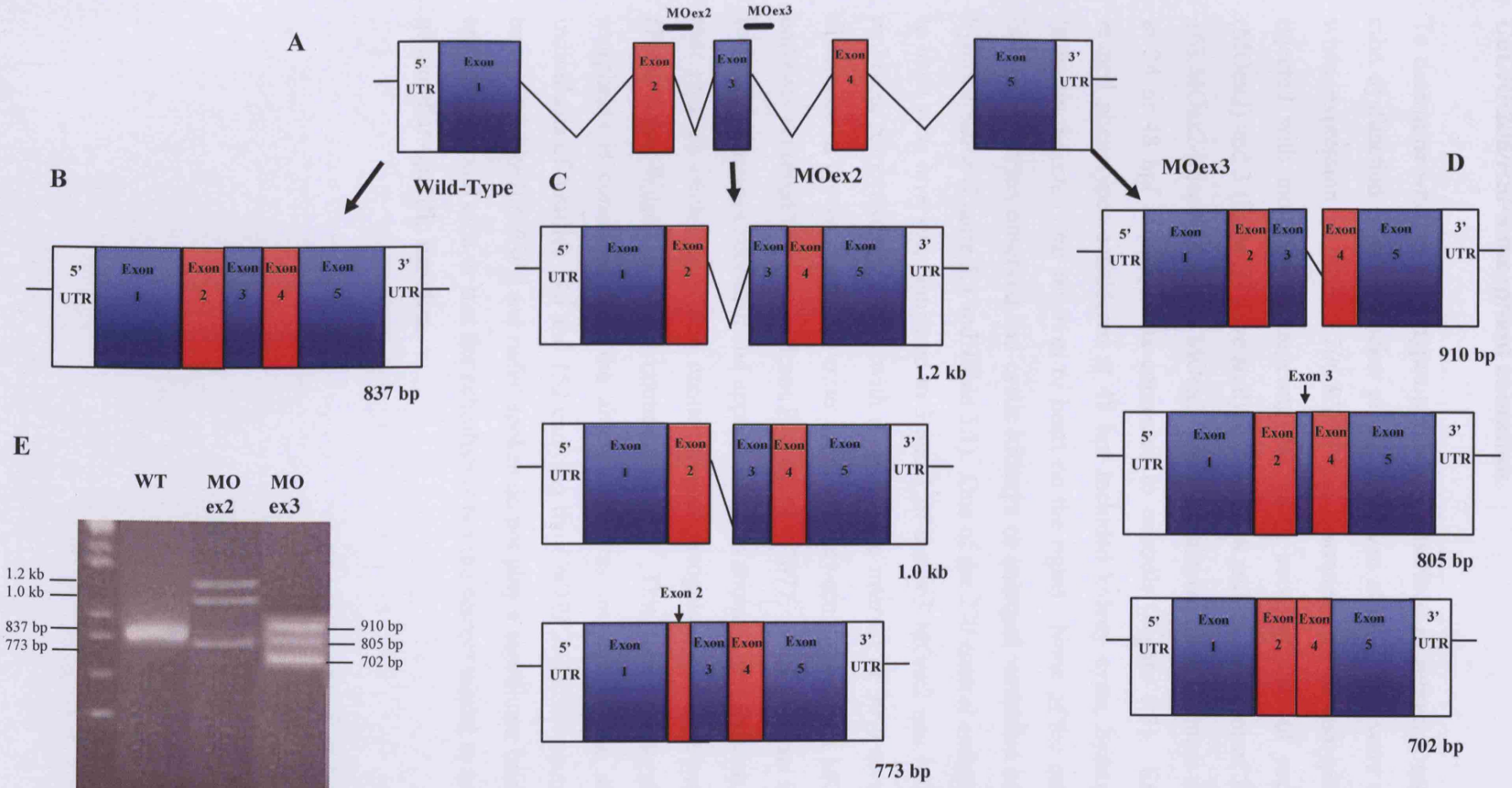


Figure 5.2: Morpholino effects on splicing of zebrafish *C6ORF206/RSP9*. **A:** Genomic zebrafish *C6ORF206/RSP9* (7.2 kb) **B:** Structure of the 5 exon wild-type zebrafish *C6ORF206/RSP9* transcript (837 bp) confirmed by RT-PCR and sequencing. **C:** Morpholino MOex2 disrupts splicing of zebrafish *C6ORF206/RSP9* - three different products of 1.2 kb, 1.0 kb and 773 bp were detected by RT-PCR and sequencing. **D:** Morpholino MOex3 disrupts splicing of zebrafish *C6ORF206/RSP9* - three different products of 910 bp, 805 bp and 702 bp were detected by RT-PCR and sequencing. **E:** Products amplified from wild-type and morpholino treated zebrafish (72 hpf) showing splicing of *C6ORF206/RSP9* is disrupted by Morpholino injection.

5.1.2.2 Identification of phenotypes related to cilia dysfunction in *C6ORF206/RSP9* morphant zebrafish

To determine whether phenotypes previously described in zebrafish morphants with cilia dysfunction (Kramer-Zucker et al., 2005; van et al., 2008) were also observed when expression of *C6ORF206/RSP9* was disrupted, the phenotype of embryos injected with morpholinos targeting the splice acceptor sites of coding exons 2 (MOex2) and 3 (MOex3) were analysed at 24, 48 and 72 hpf. Out of 270 wild-type, 163 MOex2-injected and 163 MOex3-injected embryos, no differences were observed at 24 or 48 hpf in morphants compared to controls (Figure 5.3). Expected cilia-related phenotypes considered at 48 hpf included kidney cysts, hydrocephalus and laterality defects (*situs inversus* or heart on the right). None of the control or MO-injected embryos observed had cystic kidneys or enlarged ventricles consistent with hydrocephalus (Figure 5.3 and Table 5.1). One of the 270 control embryos was found to have *situs inversus*, compared to 3 and 1 of the 163 MOex2 and MOex3-injected embryos, respectively. T-tests with a confidence interval of 99% revealed that the difference in levels of *situs inversus* observed between control and MOex2-injected embryos achieved a non-significant P-value of 0.7077. *Situs inversus* is observed at low levels amongst zebrafish and appears to be an anomaly in zebrafish development that presents occasionally and creates a well recognised "background" phenotype (Prof. Steve Wilson, personal communication). This lack of laterality defect in morphants is consistent with the absence of *situs inversus* in the seven affected individuals of families 146 and 152 carrying the *C6ORF206/RSP9* mutation. It may be that *C6ORF206/RSP9* and radial spokes do not play a significant functional role in embryonic nodal cilia or that the zebrafish is not the correct model to analyse a more mammalian-specific function.

	WT	ex2	ex3
No. live embryos (n) at 48 hpf	270	163	163
No. with *situs inversus*	1	3	0
No. with cystic kidneys	0	0	0
No. with hydrocephalus	0	0	0

Table 5.1: No significant phenotype observed in splice-blocking morpholino injected zebrafish embryos at 24 or 48 hpf.

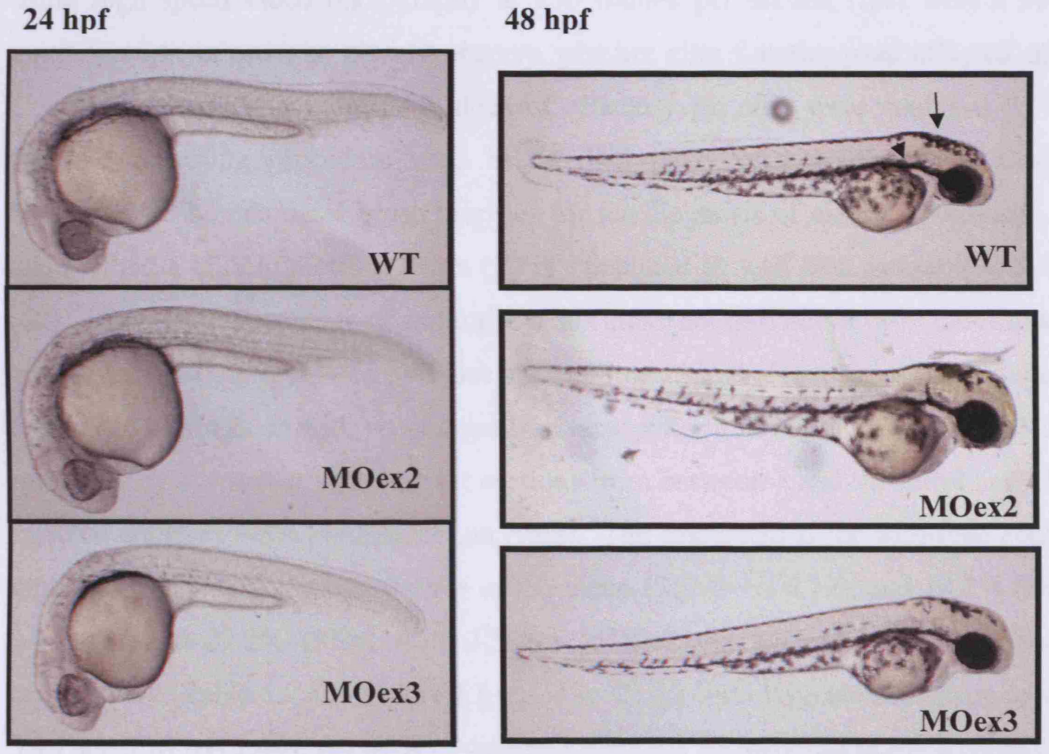


Figure 5.3: Zebrafish *C6ORF206/RSP9* knockdown morphant embryos display no obvious phenotype compared to wild type (WT). Above: Comparison of body plan between WT, MOex2 and MOex3 morphants shows no gross abnormalities. Brain ventricles (arrow) and kidney (arrow head) are indicated. 300X magnification. **Right:** Example of kidney cysts (arrow head) and hydrocephalus (arrow) seen in *DNAH9* knockdown morphants. Image taken from (Kramer-Zucker et al., 2005).

5.1.2.3 *C6ORF206/RSP9* morphants display dysmotile olfactory pit cilia compared to controls

The zebrafish olfactory pit cilia were initially observed using a high speed camera and it was observed by eye that these might be affected by cilia dysmotility. The olfactory pit cilia of the MO injected embryos appeared to have a different beat pattern to controls, although the exact nature and extent of this defect was unclear by eye. This prompted an effort to investigate the olfactory pit cilia function in more detail. To this end, in collaboration with Dr. Robert Hirst and Prof. Christopher O'Callaghan, University of Leicester, olfactory pit cilia from live control and *C6ORF206/RSP9* target MO-injected embryos at 72 hpf (Figure 5.4) were recorded using high speed video photography at 250 frames per second (fps) with a 500X magnification in order to closely observe whether cilia function was affected upon *C6ORF206/RSP9* knockdown. Videos of olfactory pit cilia were analysed by the standard operating procedure used in the University of Leicester PCD clinical laboratory on human nasal brush biopsies for the diagnosis of ciliary dyskinesia. In this method a cilia Immotility Index (II) is calculated as well as a dysmotility Index (DI). The mean percentage of immotile cilia (those not moving or only moving as a consequence of being moved by those surrounding it) and also a mean percentage of dyskinetic cilia (those with an unorganised beat pattern, twitching or immotile) was measured by eye across olfactory pit sections from between 7 and 9 control and MO-injected embryos (95% confidence intervals). The calculated II for wild-type control embryos was 1.43% (Standard error of the mean (SEM) +/- 0.12), and 18.2% (SEM +/- 0.53) and 22.2% (SEM +/- 0.32) for MOex2 and MOex3-injected embryos, respectively (Table 5.2 and Figure 5.5A). The DI for wild-type control embryos was 3.13% (SEM +/- 1.44), for MOex2-injected embryos 65.86% (SEM +/- 4.2) and for MOex3-injected embryos 65.56% (SEM +/- 2.73) (Table 5.2 and Figure 5.5B). All data were tested by one-way anova and individual data sets were compared using an unpaired Students t-test with Bonferoni correction for repeated measures. Statistical comparisons between datasets revealed that the Immotility and Dysmotility Index calculated for both sets of MO-injected embryos was significantly higher than observed for controls (p value <0.05) (Table 5.3).

5.1.2.4 *C6ORF206/RSP9* morphants display wild-type ciliary beat frequency

The average cilia beat frequency (CBF) was calculated from the same control and morphant olfactory pit cilia high-speed video data (also described in Section 2.4.19). The CBF for controls was calculated at 43.6 Hz (SEM +/- 3.72), for MOex2-injected embryos 42.32 Hz (SEM +/- 0.83) and for MOex3-injected embryos 40.03 Hz (SEM +/- 1.8) (Table 5.2). Statistical analysis revealed that there was no significant difference between the three datasets (p values >0.05) (Table 5.3). These observations are consistent with cilia beat velocity and pattern for family 146 homozygous carriers of the c.804_806delGAA mutation in *C6ORF206*, where CBF was within the normal range but the ciliary beat pattern was abnormal (Stannard et al., 2004).

5.1.2.5 *C6ORF206/RSP9* morphants display a disorganised beat pattern

Observation of zebrafish olfactory pit cilia in videos slowed down to 7 fps revealed a disorganised beat pattern (videos in Appendix 3). This resembled the circular beat pattern seen for nasal epithelial cilia samples from family 146 individuals homozygous for the *C6ORF206* c.804_806delGAA mutation (Stannard et al., 2004). It was noted in control embryos that debris was swept across the olfactory pit, as expected with a normal ciliary beat frequency and pattern in control embryos, resulting in constant clearance of debris from the pit. However, accumulation of debris at the olfactory pits was prevalent over time in MO-injected embryos with defective II and DI and a disorganised beat pattern. This failure to keep the pit clear is consistent with an ineffective beat and a vortex created by the defective circular ciliary beat pattern (Figure 5.6).

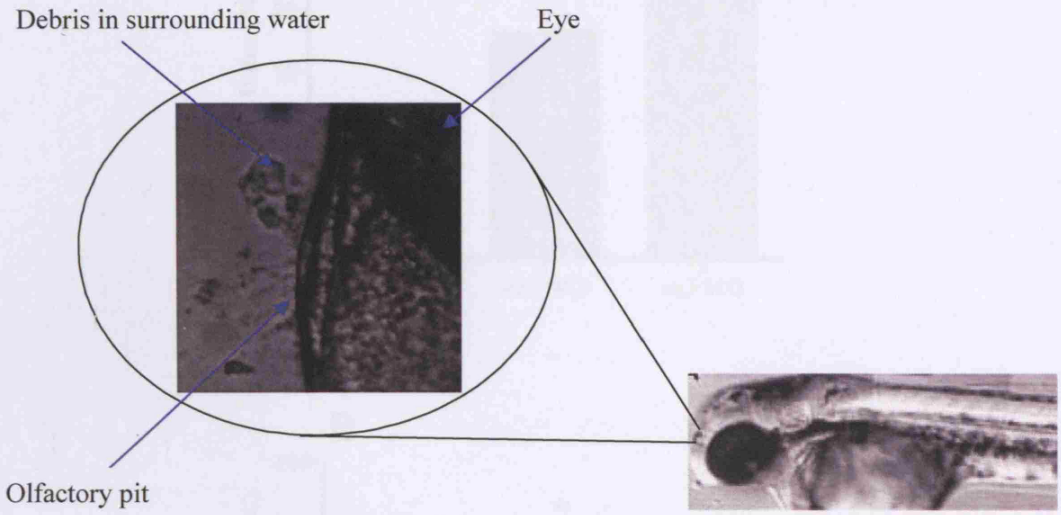


Figure 5.4: Still image of zebrafish olfactory pit observed by high speed microscopy (x100 magnification).

	WT	ex2 MO	ex3 MO
Number of embryos (n)	8	7	9
Immotility Index (% +/- SEM)	1.43 +/- 0.12	18.2 +/- 0.53	22.2 +/- 0.32
Dyskinesia Index (% +/- SEM)	3.13 +/- 1.44	65.86 +/- 4.20	65.56 +/- 2.73
CBF (Hz/bps +/- SEM)	43.6 +/- 3.72	42.32 +/- 0.83	40.03 +/- 1.8

Table 5.2: Immotility Index, Dyskinesia Index and ciliary beat frequency (CBF) of *C6ORF206/RSP9* knockdown morphants in comparison to wild-type controls.

	WT-MOex2	WT-MOex3	MOex2-MOex3
Immotility Index	P<0.0001	P<0.0001	P=3191
Dyskinesia Index	P<0.0001	P<0.0001	P=0.9511
CBF (Hz/bps)	P=0.7577	P=0.3836	P=0.3116

Table 5.3: Calculated statistical comparison within Table 5.2 datasets – t-test indicates the immotility and dysmotility indexes are significantly different in morphants compared to controls (P<0.05), whereas the CBF is not significantly different (columns 1 and 2).

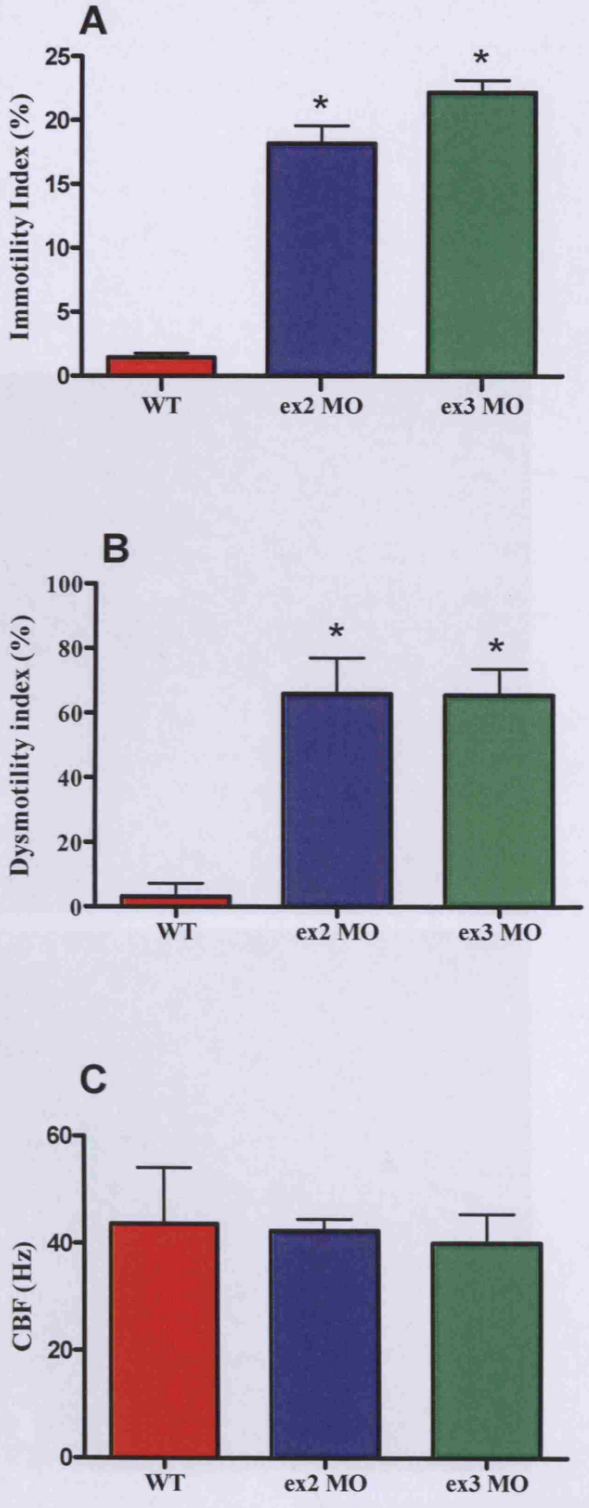


Figure 5.5: Graphical representation of data from Tables 2 and 3. Immotility index, Dysmotility Index and CBF in zebrafish *C6ORF206/RSP9* knockdown morphants compared to controls. Immotility Index and Dysmotility Index significantly higher in morphants compared to controls (asterix indicates p value <0.0001). CBF not significantly different.

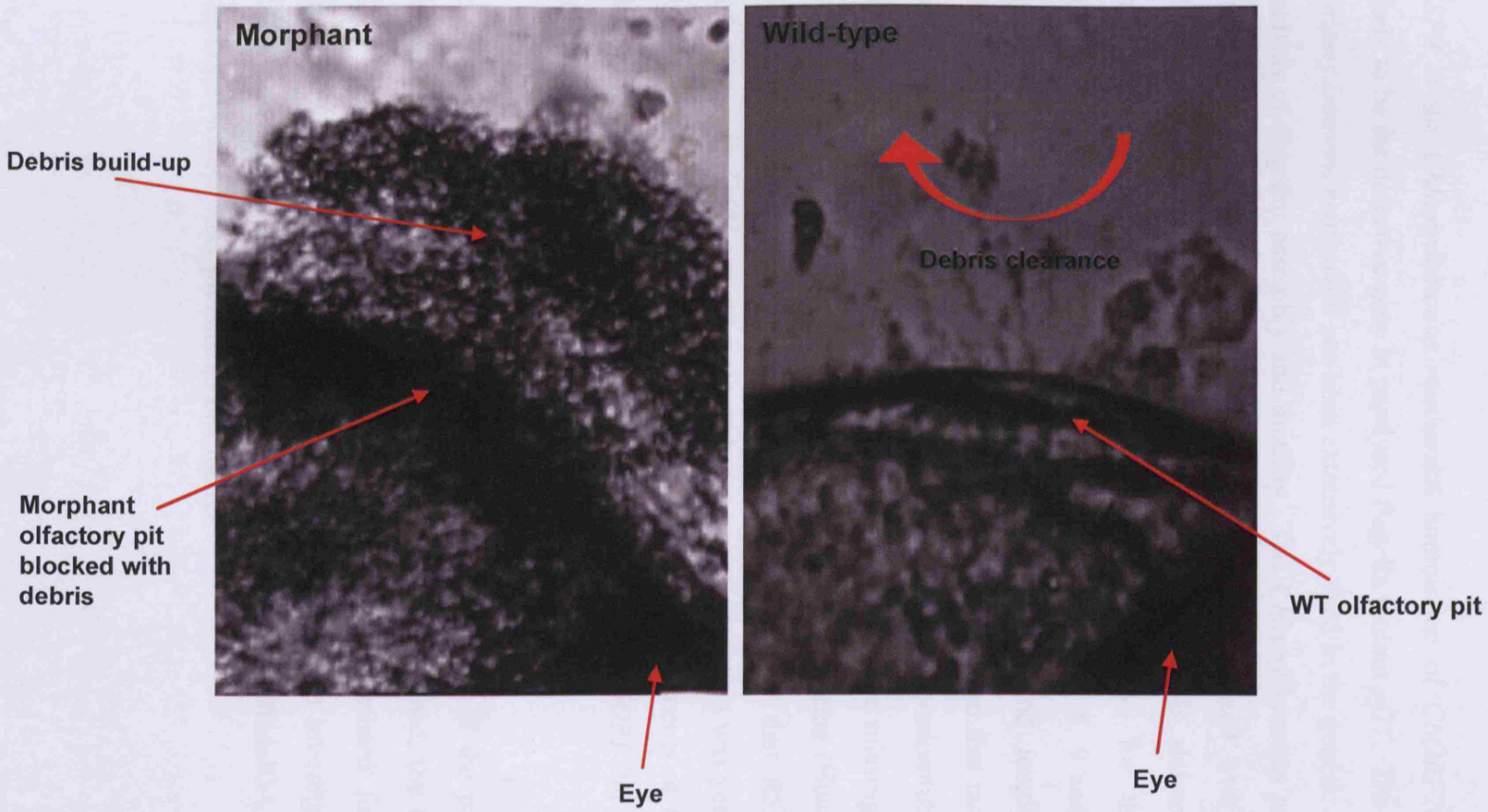


Figure 5.6: Accumulation of debris in olfactory pits of *C6ORF206/RSP9* knockdown morphant zebrafish.

5.1.3 Analysis of effects of c.804_806delGAA (K268del) mutation on C6ORF206/RSP9 function using the *Chlamydomonas reinhardtii* model

RSP9 is the *Chlamydomonas reinhardtii* homologue of *C6ORF206/RSP9* and is likely to be the defective gene in paralysed flagella mutant *pf17*. The biflagellate alga *Chlamydomonas reinhardtii* has been extensively used in the genetic and biochemical analysis of flagellar assembly and motility. The *Chlamydomonas* paralysed flagella mutant *pf17* was created by UV irradiation and isolated over 30 years ago (EBERSOLD et al., 1962). Protein biochemical analysis showed that it has a molecular deficiency, demonstrated by autoradiography, for the five protein components of the axonemal radial spoke head, RSP1, 4, 6, 9 and 10 (Figure 1.7 Section 1.1.4.2.2) (Yang et al., 2006). Transverse and longitudinal electron microscopy images of isolated axonemes from *pf17* revealed that radial spoke stalks extend with the same periodicities as wild-type from outer-microtubule A, however the stalks are disorganized and the radial spoke heads were missing and the central pair was seen to be displaced toward one side of the axoneme (Huang et al., 1981). Dikaryon rescue experiments provided indirect evidence that *RSP9* may be the defective gene product in this mutant but the exact mutation was yet to be identified (Huang et al., 1981). A study of the radial spoke components of *Chlamydomonas* flagella published during the course of this study found *C6ORF206* to be the human homologue of *Chlamydomonas RSP9* (Yang et al., 2006).

In order to prove that a mutation in *RSP9* was the cause of the paralysed flagella phenotype in *pf17* which had never been directly established, the *Chlamydomonas RSP9* gene from *pf17* mutant strain CC-1035 was screened for mutations by sequencing. This was to establish its utility as a model to investigate the effect of *RSP9* mutations mimicking the human *C6ORF206* c.804_806delGAA mutation, and functional studies were pursued in this model.

5.1.4 Mutational Analysis of *RSP9* reveals a single base pair deletion in *pf17*

Sequencing of the five coding exons, intron-exon boundaries and 40 bp into the 5' and 3' UTRs of *RSP9* in genomic DNA isolated from wild-type and mutant *pf17* *Chlamydomonas* revealed a single base pair deletion in exon 2 (131delG) present in *pf17* but not wild type (Figure 5.7A). The predicted effect of this mutation on the RSP9 protein product is a frameshift after residue S45 and a severe truncation of the 269 amino acid protein (Figure 5.7C), likely resulting in a null mutation as the 45 amino acid protein remaining would not be expected to retain any function or might be unstable and degraded.

5.1.4.1 Restriction digest shows the RSP9 131delG mutation is present in *Chlamydomonas pf17* progeny strains

The *RSP9* 131delG mutation identified in genomic DNA from *Chlamydomonas pf17* destroys *BspEI* restriction enzyme site (Figure 5.7A and 5.8). To demonstrate that the mutation identified had not arisen since isolation of *pf17* in the 1970's, but was the true original mutation responsible for the paralysed flagella phenotype, restriction digestion using *BspEI* was performed on a portion of *RSP9* amplified by PCR from genomic DNA (primer sequences in Appendix 3). DNA isolated from the original *pf17* mutant strain CC-1035, plus four *pf17* progeny strains: CC-1332, CC-1143, CC-2645 and CC-262 and wild-type CC-1732 was amplified and digested with *BspEI* restriction enzyme. The undigested product gave rise to a single 1.1 kb band in agarose gel electrophoresis, whereas *BspEI* digested wild-type *RSP9* PCR product was present as three bands of 636 bp, 279 bp and 194 bp. The *Chlamydomonas* strains with the 131delG mutation yielded two bands of 830 bp and 279 bp (Figure 5.7B and 5.8). Thus, the 131delG mutation identified in mutant *pf17* was present in progeny strains and not the wild-type control, confirming that this deletion was most likely to be the original *pf17* mutation causing the paralysed flagella phenotype.

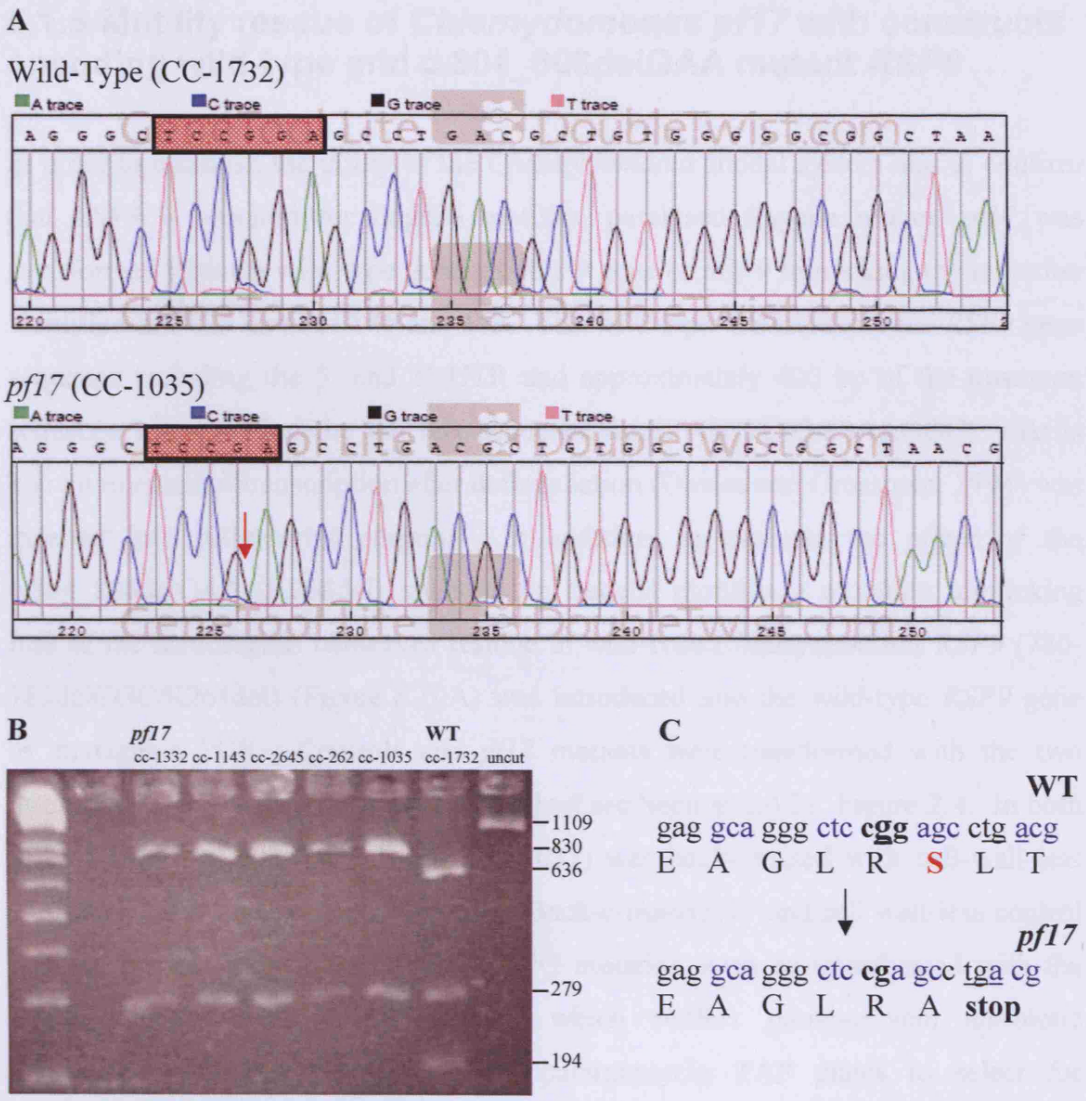


Figure 5.7: Mutation revealed in *RSP9* in *Chlamydomonas* paralysed flagella strain *pf17*. A: Single base pair deletion (131delG) in *RSP9* revealed in *pf17 Chlamydomonas reinhardtii* mutant that was not present in control, indicated by red arrow. Mutation destroys *BspEI* restriction enzyme site (highlighted by red box). **B:** *BspEI* restriction digest of amplified *RSP9* PCR products from *Chlamydomonas reinhardtii pf17* progeny strains (CC-1332, CC-1143, CC-2645 and CC-262) compared to WT (CC-1732) and undigested product. Presence of 131delG mutation yields restriction digest products of 830 bp and 279 bp. Absence of the 131delG mutation results in products of 636 bp, 279 bp and 194 bp. The 131delG mutation is therefore present in progeny *pf17* strains, indicating that this is the true *RSP9* mutation causing the paralysed flagella phenotype. **C:** 131delG mutation predicted to cause a frameshift mutation after residue S45 (red), resulting in a severe truncation of the 269 amino acid protein.

5.1.5 Motility rescue of *Chlamydomonas pf17* with constructs encoding wild-type and c.804_806delGAA mutant *RSP9*

In order to establish the utility of the *Chlamydomonas* model system and to confirm that *RSP9* is required for flagella motility, paralysed flagella mutant *pf17* was transformed with the wild-type genomic *RSP9* gene or *RSP9* mimicking the mutation identified in PCD families 146 and 152. The wild type *Chlamydomonas RSP9* gene sequence including the 5' and 3' UTR and approximately 400 bp of the upstream sequence incorporating the tub box sequence motif (GCTC(G/C)AAGGC) that is known to enhance transcription after deflagellation (Davies and Grossman, 1994) was inserted into pBluescript plasmid. In addition, to evaluate the effect of the c.804_806delGAA (K268del) mutation on flagella motility, a mutation mimicking loss of the homologous conserved residue in wild-type *Chlamydomonas RSP9* (780-783delCGC/R261del) (Figure 5.10A) was introduced into the wild-type *RSP9* gene by mutagenic PCR. Controls and *pf17* mutants were transformed with the two resulting constructs. For mutagenesis method see Section 2.4.21, Figure 2.4. In both cases *Chlamydomonas pf17* (strain CC-1035) was back-crossed with cell-wall-less mutant *cw15*, for ease of transformation. Back-crossed *pf17* and cell wall-less control cultures (*cw15*) without the *RSP9* 131delG mutation were co-transformed with the *RSP9* construct and plasmid pSI103, which confers paramomycin antibiotic resistance. Colonies were grown on paramomycin TAP plates to select for transformants. Ninety-seven wild-type and fifty-six mutant *RSP9* transformed colonies were grown and picked for DNA extraction. Only one colony from each transformed set was a successful transformant and this was maintained for future motility analysis. For the purposes of this study, the cell-wall-less control is referred to as *cw15* (wild-type *RSP9*). The mutant *pf17* strain transformed with wild-type *RSP9* or mutant *RSP9* construct is referred to as *pf17*-T and *pf17*-Tmut, respectively.

5.1.5.1 Confirmation of incorporation of introduced DNA in *Chlamydomonas* transformant strains

After transformation, genomic DNA from *Chlamydomonas* controls, mutant *pf17* and *pf17 RSP9* transformants was analysed by restriction digest to confirm incorporation of the wild-type and mutagenic *RSP9* constructs into the genetic material of the transformed strain. The original *pf17* strain *RSP9* 131delG mutation destroys a *BspEI*

restriction site (Figure 5.8). Digestion of the *RSP9* PCR product from the cell wall-less non-transformed *pf17* strain with *BspEI* restriction enzyme gives two bands of 830 bp and 279 bp. Digestion of the *RSP9* PCR product from the control *cw15* strain carrying wild-type *RSP9* gives three bands of 636 bp, 279 bp and 194 bp. The *pf17*-T strain yielded four bands of 830 bp, 636 bp, 279 bp and 194 bp. This confirmed the presence of both wild-type and 131delG *RSP9* in this transformant (Figure 5.9).

The c.780_783delCGCmutation introduced into *Chlamydomonas* to generate the *pf17*-Tmut strain is equivalent to the c.804_806delGAA in human PCD families 146 and 152, since the amino acid residue at position 261 is deleted (Chapter 4). Introduction of the wild-type *RSP9* construct into the *Chlamydomonas* genome results in the strain carrying the endogenous 131delG *RSP9* gene, plus wild-type *RSP9*. Introduction of the human-equivalent mutation into the *Chlamydomonas* genome results in the strain carrying the endogenous 131delG *RSP9* gene, plus the c.780_783delCGCmutated copy. The c.780_783delCGCmutation destroys an *FspI* restriction enzyme site. Digestion of *RSP9* PCR product from *cw15* control, *pf17* and *pf17*-T with *FspI* therefore allows identification of correct incorporation of the mutant gene. *FspI* digestion gives four bands of 1.8 Kb, 495 bp, 337 bp and 268 bp due to absence of c.780_783delCGC (Figure 5.10C – only upper bands shown). Digestion of PCR product from the *pf17*-Tmut strain gives five bands of 2.3 Kb, 1.8 Kb, 495 bp, 337 bp and 268 bp due to the presence of wild-type and mutant *RSP9*, arising from the lack of an *FspI* restriction site in the mutagenic form of *RSP9*. This confirmed the presence of both wild-type and mutagenic c.780_783delCGC*RSP9* in this transformant (Figure 5.10C – only upper two bands shown).

These results showed that both the wild type and mutant *RSP9* construct incorporated correctly into the genetic material of the *Chlamydomonas reinhardtii* mutant *pf17*–T and *pf17*–Tmut strains, respectively.

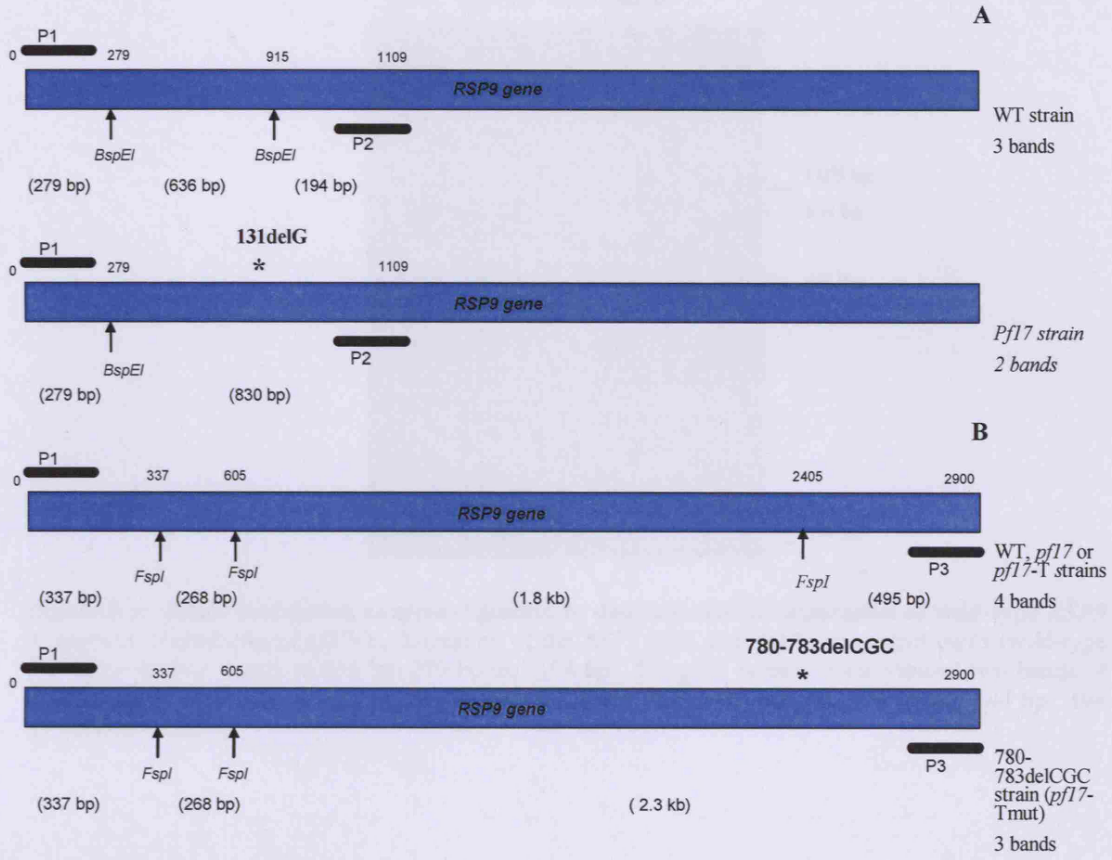


Figure 5.8: *BspEI* and *FspI* restriction sites in *Chlamydomonas RSP9* are destroyed, respectively, by the *pf17* 131delG mutation and the c.780_783delCGCmutation that mimics the human c.804_806delGAA mutation. A) *BspEI* restriction enzyme sites indicated with arrows, one of which is destroyed by the 131delG mutation. Primers highlighted with black blocks. B) *FspI* restriction sites indicated with arrows, one of which is destroyed by the c.780_783delCGCmutation. Note, *pf17*-Tmut has the c.780_783delCGC*RSP9* gene as well as the wild-type bands. Primers highlighted with black blocks (reverse primers for A and B used were positioned at different locations within the gene in order to include the relevant mutation. Numbers above the gene indicate restriction enzyme site and primer positions, numbers in brackets below the gene indicate restriction fragment sizes after digestion.

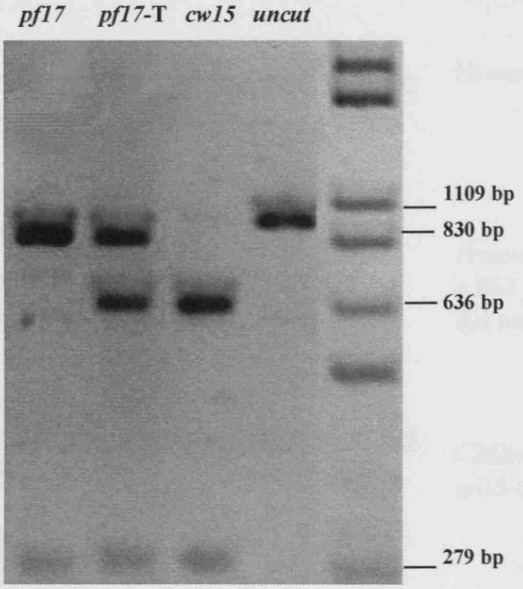


Figure 5.9: *BspEI* restriction enzyme digestion to demonstrate incorporation of wild-type *RSP9* in rescued transformant *pf17*-T. Digestion of the *RSP9* PCR product from control *cw15* (wild-type *RSP9*) gives three bands of 636 bp, 279 bp and 194 bp. The *pf17* mutant strain yielded two bands of 830 bp and 279 bp. The *pf17*-T strain yielded four bands of 830 bp, 636 bp, 279 bp and 194 bp. 194 bp band not shown.

C

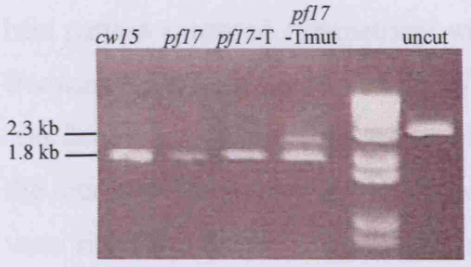


Figure 5.10: (A) The family 146 and 152 c.804_806delGAA mutation results in deletion of residue K268 in the human C6ORF206 protein. The 3' end of the gene and the C-terminal of the protein are shown. **(B) The c.780_783delCGCmutation introduced into wild-type *Chlamydomonas RSP9* gene (homologue of human *C6ORF206/RSP9*) results in deletion of conserved R261 in *Chlamydomonas*.** The 3' end of the gene and the C-terminal of the protein are shown. Small black arrow indicates *FspI* restriction enzyme site CTGCGC, that is lost upon introduction of the c.780_783delCGCmutation. **(C) *FspI* restriction enzyme digestion to demonstrate incorporation of mutant *RSP9* in part-rescued transformant *pf17*-Tmut.** Digestion of *RSP9* PCR product from *cw15* control, *pf17* and *pf17*-T with *FspI* gives four bands of 1.8 Kb, 495 bp, 337 bp and 268. Digestion of PCR product from the *pf17*-Tmut strain gives five bands of 2.3 Kb, 1.8 Kb, 495 bp, 337 bp and 268 bp. *pf17*-Tmut carries both wild-type and mutant *RSP9*. 495 bp, 337 bp and 268 bp bands are not shown.

5.1.5.2 Flagella dysmotility phenotype of *Chlamydomonas* transformants

pf17 colonies transformed with wild-type or mutant *RSP9* were cultured and viewed under a microscope. Out of 97 *pf17 RSP9* transformed (*pf17*-T) colonies, one appeared to have regained motility similar to wild type *Chlamydomonas.* Out of 56 c.780_783delCGCmutant *RSP9* transformed (*pf17*-Tmut) colonies, one appeared to have regained motility but the transformants were moving at about half the speed of wild type as judged by eye under might microscopy. This prompted a more detailed high speed video analysis.

Four different strains: control *cw15*, mutant *pf17*, transformant *pf17 Chlamydomonas* cultures transformed with wild-type *RSP9* (*pf17*-T) and transformed *pf17* cultures with mutagenic *RSP9* (*pf17*-Tmut) were viewed under a microscope under controlled humidity (80%) and temperature (30°C) and videos were recorded using a high speed camera (500 frames per second (fps)). The *Chlamydomonas* cultures included cells at various stages of the cell cycle, including flagella-less forms, so only cells where the beat pattern appeared symmetrical were used in the analysis. The mean flagella beat frequency (FBF) of 20-26 individual cells was calculated and their Immotility Index was determined by counting the number of cells with static flagella as a percentage of the total number present in 5-10 slowed down video sequences. These measurements were similar to those done in Section 5.1.2.3 on zebrafish olfactory pit cilia. All data were tested by one-way anova and individual data sets were compared using an unpaired Students t-test with Bonferoni correction for repeated measures. This analysis was carried out by Dr. Robert Hirst, Department of Microbiology and Immunology, University of Leicester.

The calculated FBF was 47.9 Hz for control *cw15*, 0 Hz for static flagella mutant *pf17*, 53.5 Hz for *pf17*-T and 24.9 Hz for *pf17*-Tmut (Table 5.4 and Figure 5.11). There was no significant difference in the FBF of *cw15* (control) and the *pf17*-T strain rescued with wild type *RSP9*. The *pf17* strain flagella were all static. There was a significant ($p<0.05$) decrease in the FBF of *pf17*-Tmut compared with the *cw15* and pf17-T strains, to approximately half their values.

The Immotility Index was determined to be 1% in control *cw15*, 99.8% in *pf17*, 4.6% in *pf17*-T and 73% in *pf17*-Tmut (Table 5.5 and Figure 5.11). There was a significant difference in the Immotility Index between *cw15* control and *pf17* (99% increase), where the latter were all static and *cw15* were motile at expected wild type levels. The *pf17*-T rescued strain was in line with control *cw15* for flagella immotility and was similarly significantly different to *pf17*, showing a decrease of 95.2%. The *pf17*-Tmut strain had a significantly different Immotility Index compared to both *cw15* control and *pf17* ($p<0.05$): *cw15* and *pf17*-T had almost full motility and *pf17* was fully immotile, whereas the *pf17*-Tmut strain had 73% immotile flagella, reflecting a smaller reduction in motility compared to the putative null allele causing static flagella in *pf17*.

This data indicated that introduction of the wild-type *RSP9* gene onto an RSP9-null background (*pf17*) rescues flagella motility and that introduction of *RSP9* carrying the c.780_783delCGCmutation (equivalent to c.804_806delGAA (K268del) in human PCD families 146 and 152 – Chapter 4) restores only a proportion of motility. Therefore the *C6ORF206/RSP9* homologue in *Chlamydomonas reinhardtii*, *RSP9*, is essential for flagella function and the mutation identified in families 146 and 152 could indeed cause the reduced motility observed in their respiratory cilia. It was not possible to make any conclusions about changes to the waveform or its coordinated action in the *Chlamydomonas* model, due to the variability of motility amongst *Chlamydomonas* in these non-synchronous cultures.

	cw15	*pf17*	*pf17-T*	*pf17-Tmut*
Number of cells (n)	21	26	20	21
FBF (Hz/bps +/- SEM)	47.9 +/- 3.0	0.0 +/- 0.0	53.5 +/- 3.13	24.9 +/- 2.43

Table 5.4: Flagella Beat Frequency (FBF) in cell wall-less *RSP9* transformants. *cw15* is the motile control, *pf17* (cell wall-less) is the static flagella strain, *pf17*-T is the wild-type *RSP9* transformant of *pf17*, *pf17*-Tmut is the mutagenic *RSP9* (780-783delCGC) transformant of *pf17*. Data represents the means and SEM of 20-26 cells per strain.

	cw15	*pf17*	*pf17-T*	*pf17-Tmut*
Number of recordings (n)	10	5	7	10
Immotility Index (% +/- SEM)	1.0 +/- 0.66	99.8 +/- 0.20	4.6 +/- 1.4	73 +/- 5.22

Table 5.5: Immotility Index in *RSP9* transformants. *cw15* is the motile control, *pf17* (cell wall-less) is the static flagella strain, *pf17*-T is the wild-type *RSP9* transformant of *pf17*, *pf17*-Tmut is the mutagenic *RSP9* (780-783delCGC) transformant of *pf17*. Data represents the means and SEM of 5-10 video recordings per strain.

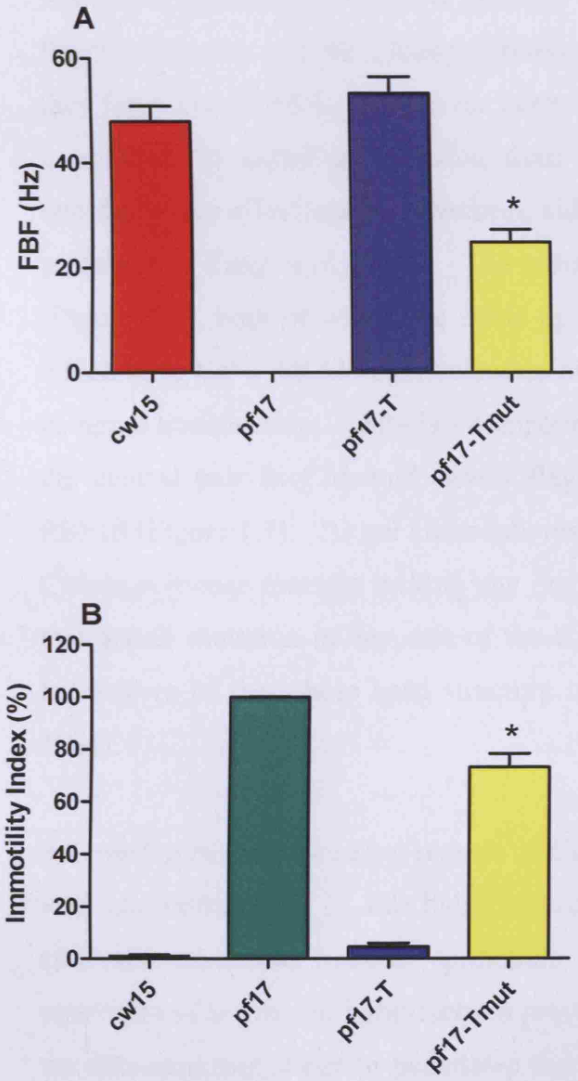


Figure 5.11: Graphical representation of data in tables 5.4 and 5.5. The mean flagella beat frequency (FBF) (A) and mean Immotility Index (B) of four strains of *Chlamydomonas*. * indicates p value <0.001 – *pf17*-Tmut is significantly different from *cw15*, *pf17* and *pf17*-T.

5.2 *Discussion*

Following the identification in chapter four of a homozygous 3 bp deletion (c.804_806delGAA) in *C6ORF206/RSP9* in DNA from all seven affected individuals of two Arabic families with PCD, functional analysis was performed to investigate the *C6ORF206/RSP9* gene and to determine if the single *RSP9* mutation found in the UCL PCD family collection was pathogenic. The deletion is predicted to cause an in-frame single conserved amino acid deletion (K268del) in the C-terminus of the C6ORF206/RSP9 protein. Although the 3 bp deletion was not present in 126 population matched control chromosomes, this may not be sufficient to provide definitive proof that this is not a non-pathogenic polymorphism. In order to ensure a 95 % confidence that this is not a polymorphism with a frequency of 0.01, 340 normal control chromosomes should be analysed (Collins and Schwartz, 2002).

The radial spokes are located at periodic intervals along the length of the cilia and flagella axoneme. In the *Chlamydomonas* model system it has been established that they form a connection between the central pair and peripheral microtubules to act as a scaffold for signal transduction from one component to the other in order to coordinate the cilia/flagella waveform, although the precise mode of action is poorly understood (Yang et al., 2006). The radial spoke consists of a "stalk" and a "head" (Figure 1.7), both of which are made up of many protein subunits, a proportion of which have Ca^{2+}, AKAP and nucleotide-binding domains that are thought to function in signal transduction. RSP9 is a component of the radial spoke head located close to the central pair in *Chlamydomonas* flagella, along with RSP1, RSP4, RSP6 and RSP10 (Figure 1.7). 2D gel electrophoresis of axonemal polypeptides prepared from *Chlamydomonas* mutants lacking any one of the radial spoke head proteins revealed that a null mutation in any one of the five radial spoke head components leads to breakdown of the whole head structure in the axoneme (Cole, 2005; Huang et al., 1981).

An overlap in the expression pattern of *C6ORF206/RSP9* and the PCD gene *DNAH5* was demonstrated by *in situ* hybridisation in mouse tissues known to carry motile cilia: the nasal and tracheal epithelium and the embryonic node. Although the resolution of the *in situ* hybridization prevented analysis of expression specifically in the cilia axoneme, it can be postulated that *C6ORF206/RSP9* is expressed specifically

in ciliated cells due to its homology with the *Chlamydomonas RSP9* gene, which encodes a radial spoke head protein (Yang et al., 2006). *C6ORF206/RSP9* expression was not, however, expected in the embryonic node due to the absence of *situs inversus* in affected individuals of families 146 and 152. It had not previously been established whether radial spokes are present in nodal cilia, although it is known that the nodal cilia have a 9+0 structure, i.e. no central pair. If radial spokes are present in nodal cilia as suggested by node-specific staining of *RSP9* they may have a redundant or reduced function as there is no central pair to relay signals from to the inner dynein arms on the peripheral microtubule doublets. It is possible that rather than play a direct role in motility as they do in 9 + 2 cilia, they could be present to provide structural integrity to the 9 + 0 axoneme.

The ultrastructural defect caused by the c.804_806delGAA (K268del) mutation in family 146 and perhaps also family 152 is an intermittent absence of the central pair, giving rise to a normal ciliary beat frequency (CBF) but with a circular beat pattern that is not effective for mucociliary clearance (Stannard et al., 2004). Nodal cilia are thought to have a circular beat pattern, so it may be consistent that any defect resulting in a circular beat pattern for respiratory cilia would not cause a defect in nodal cilia. This was in fact previously discussed by Stannard *et al* (Stannard et al., 2004). Thus, *C6ORF206/RSP9* is present at the mouse embryonic node but its exact purpose is not yet clear. It could even be residual and serve no specific function.

Following *C6ORF206/RSP9* gene knockdown in zebrafish by injection of splice-targeting morpholino oligonucleotides, no immediately obvious cilia defect phenotype was observed: *situs inversus*, hydrocephalus or cystic kidneys. However, after closer observation of olfactory pit cilia in embryos 72 hours post fertilization (hpf), it was revealed that morphants had a significantly increased olfactory cilia dysmotility and immotility index compared to wild type controls. In addition, the morphants had a normal CBF, which was at wild type levels. This phenotype was consistent with that seen in respiratory epithelial cilia from affected individuals of family 146 who have the c.804_806delGAA mutation in *C6ORF206/RSP9*. This provides support for the hypothesis that loss-of-function defects in this gene cause this unusual ciliary phenotype.

Zebrafish have a conserved nodal cilia function equivalent to human embryos and therefore provide a good model for left-right axis determination (Essner et al., 2002). The absence of *situs inversus* in the zebrafish *C6ORF206/RSP9* morphants supported the hypothesis that *C6ORF206/RSP9* does not function in the left-right axis determination role of nodal cilia. However, it should be noted that expression studies must still be performed to prove that *C6ORF206/RSP9* is in fact expressed at the Kupffers vesicle, the zebrafish equivalent of the human embryonic node. Further work could include tissue-specific RT-PCR and immunofluorescence experiments to prove protein knockdown and/or mis-localisation in the morphants. In addition, in order to provide an important control, wild-type *RSP9* mRNA would be expected to rescue the normal phenotype in a morphant and this experiment is currently underway.

RSP9 is the *Chlamydomonas reinhardtii* homologue of human *C6ORF206* and was proven in this study to be the defective gene in the *Chlamydomonas* paralysed flagella mutant strain, *pf17*. Sequencing revealed that a single base pair deletion in exon 2 of *RSP9* (131delG) was present in *pf17* but not wild type *Chlamydomonas*, which was predicted to result in a severely truncated protein and therefore produce a null background. In this chapter, the *Chlamydomonas* mutant *pf17* was significantly rescued for motility by successful transformation of the wild-type *RSP9* gene and "semi-motility" was observed in *pf17* after transformation of the *RSP9* gene containing a mutation expected to mimic the loss of the homologous amino acid that was deleted in families 146 and 152. In *Chlamydomonas*, the homologous amino acid deleted in order to mimic the K268del mutation was an arginine instead of a lysine at this position. This residue was not therefore conserved although both these amino acids have similar basic, polar properties, which suggests a possible functional conservation of this position between species rather than absolute amino acid conservation. Therefore, it was proposed that it may be the absence of the amino acid at that position, rather than loss of its specific chemical properties that may confer protein dysfunction. The *Chlamydomonas* results further proved that defects in the *RSP9* gene cause dysmotile cilia and flagella. In addition, they suggest that the in-frame single amino acid deletion identified in PCD patients causes a reduced motility compared to wild-type, a cilia dysfunction sufficient to cause disease. However, the effect of different transgene copy numbers cannot be excluded as a cause of differing

levels of motility in the *Chlamydomonas* transformants. Therefore, quantitative PCR should be performed on transformant DNA in order to confirm that copy numbers are not significantly different.

The predicted effect of the 131delG mutation identified in the *pf17* mutant *Chlamydomonas* on the RSP9 protein product is a frameshift after residue S45 and a severe truncation of the 269 amino acid protein. This likely results in a null mutation, since the 45 amino acid mutant protein translated would not be expected to retain any function and might likely be unstable and degraded. The phenotypic effect of this null mutation in *pf17* is paralysed flagella which at the ultrastructural level display an absence of radial spoke heads, disorganised spokes and displacement of the central pair towards one side of the axoneme (Huang et al., 1981). The c.804_806delGAA mutation revealed in the human *C6ORF206* gene however, was a mutation at the 3' end of the gene only 22 bp from the stop codon. The predicted effect on the RSP9 protein, the in-frame deletion of one amino acid at the C-terminus, seems likely to be milder in its deleterious effect on protein function. The phenotypic effect of dysmotile cilia, described as having a circular beat pattern and an intermittent absence of the central pair along the axoneme may therefore reflect a less disturbed RSP9 protein than the severely truncated mutant protein in the *pf17 Chlamydomonas* which have totally static flagella. Unfortunately, it was not possible to visualise the radial spokes in family 146 in EM micrographs from Prof. O'Callaghan for comparison. Table 5.6 shows a summary to compare the different mutations and phenotypes in human versus the *Chlamydomonas RSP9* mutant.

To speculate on the significance of the c.804_806delGAA human mutation, the central pair-radial spoke interactions are thought to propagate the cilia bend and direction and the radial spoke-inner dynein arm interactions are thought to play a role in the velocity of the bend (Figure 5.12) (Smith, 2002; Smith and Lefebvre, 1997). As *pf17* mutants have an absence of radial spoke heads, the signal for bend propagation could be lost and therefore the paralysed flagella phenotype presents. We propose that introduction of wild-type RSP9 as observed in the rescue experiments in this chapter, results in re-building of the radial spoke head so that the bend propagation signal is regained and flagella motility is restored. As the affected individuals of family 146 and 152 have only a slight change in RSP9 (in frame single

amino acid deletion), it can be postulated that the radial spoke head is still formed and that there is enough of an interaction with the central pair to propagate a bend, but not enough to regulate waveform, thus the CBF is normal but dysmotility and immotility is increased. This hypothesis is summarised in Figure 5.12.

In future work it would be of interest to determine the protein composition of the radial spokes in the *Chlamydomonas* strains that were created mimicking the human mutation, versus wild-type. We speculate that unlike *pf17* where all five spoke head proteins are absent from the axoneme, the *Chlamydomonas* strain mimicking the human mutation would be associated with a more preserved spoke head with a preserved but partial function. Western blots on isolated flagella axonemes using radial spoke head antisera would be a good way to approach this.

Phenotype	Human RSP9 804-806delGAA	*Chlamydomonas pf17* RSP9 131delG
RSP9 protein	Missense mutation	Null mutation
Central pair	Intermittent	Complete but displaced
Radial spoke	Not possible to visualise	Absent radial spoke heads
Cilia/flagella beat	Rotational, dysmotile	Paralysed flagella

Table 5.6: Comparison of human and *Chlamydomonas pf17 RSP9* mutations and phenotypic effects.

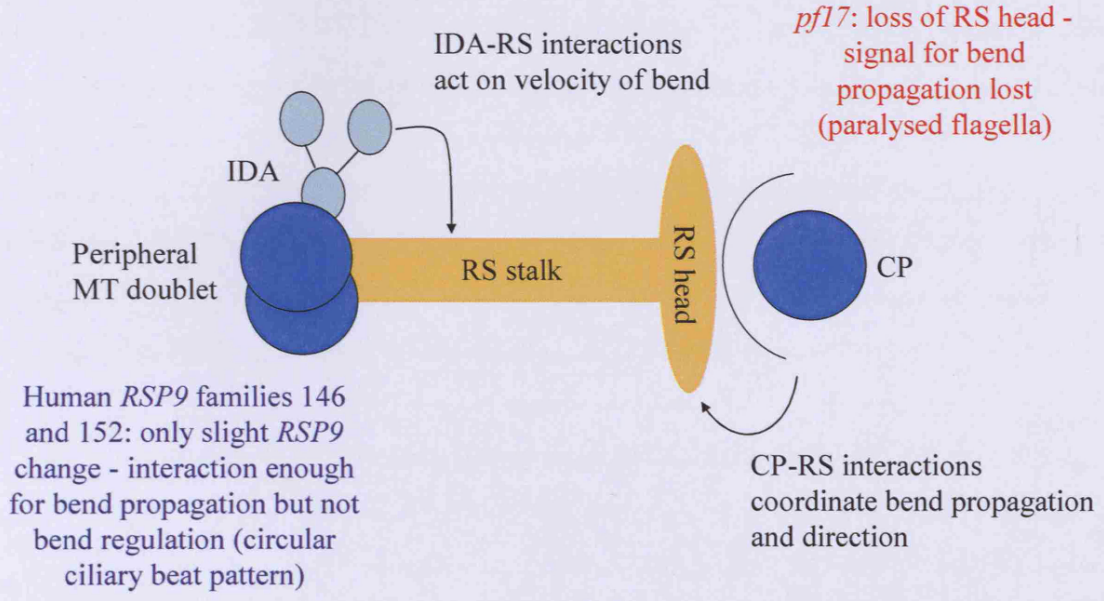


Figure 5.12: Radial spoke function and the proposed effects of RSP9 mutations on the beat pattern of cilia and flagella. RS = radial spoke, CP = central pair, IDA = inner dynein arm, MT = microtubule. Red text = effect of *pf17 RSP9* 131delG mutation. Blue text = effect of human *RSP9* c.804_806delGAA mutation.

6 CHAPTER SIX

6.1 *GENERAL DISCUSSION*

6.1.1 Conclusions from this study

This study involved genetic analysis in consanguineous PCD families using whole-genome linkage analysis by homozygosity mapping. After linkage analysis, a positional cloning approach was followed by identification of positional candidate genes and DNA sequencing in linked patients, versus controls. The results of this study revealed linkage in four out of five Pakistani families with absent inner and outer dynein arms to either chromosome 11q23.3-q24.3 or chromosome 17q21.31-q22. A comparative bioinformatic approach identified four positional candidate genes, each of which have been screened for mutations but none found as yet.

Furthermore, a locus on chromosome 6p21.2-21.1 was consistent for linkage in a large family pedigree 152 of Arabic Bedouin origin with an unknown ultrastructural phenotype. This locus overlapped with one identified in a separate genome screen in family 146 from the United Arab Emirates (UAE) with an intermittent central pair. A common mutation (c.804_806delGAA/K268del) was revealed for both families in the human homologue of *Chlamydomonas* radial spoke head protein, RSP9. *RSP9* expression was confirmed by *in situ* hybridisation in specific ciliated tissues in mouse, including the nasal epithelium, trachea and the embryonic node.

RSP9 was proven to cause cilia motility defects by complete knockdown in zebrafish morphants. A null mutation in the *RSP9* gene in pre-existing *Chlamydomonas* mutant, *pf17*, was found to underlie its paralysed flagella phenotype and this was rescued back to normal motility by incorporation of the wild-type gene into the genetic material of the alga. Introduction of the *RSP9* gene carrying the equivalent of the K268del mutation into *pf17* resulted in restoration of motility but not to wild-type levels, which is consistent with the clinical phenotype observed for family 146. These data confirm that *RSP9* is a PCD-causing gene and genes for radial spoke proteins can be considered good candidates for PCD with radial spoke and central pair defects.

This study revealed that *C6ORF206/RSP9* is expressed at the embryonic node in mouse. However, the K268del mutation did not give rise to *situs inversus* in any of seven PCD patients who were homozygous carriers of the mutation, nor did a significant number (3 out of 326) of zebrafish *RSP9* morphant embryos display laterality defects, despite complete *C6ORF206/RSP9* gene knock down. In both zebrafish morphants and human PCD patients, cilia expressing the K268del mutation displayed a dysmotile circular beat pattern which was ineffective for mucociliary clearance but that could be proposed to be similar to that observed for cilia at the embryonic node. We therefore postulate that although C6ORF206/RSP9 and presumably radial spokes are present at the node during development, they do not function to influence nodal cilia motility; and that defects in this protein do not affect the circular beat pattern required of nodal cilia for leftward nodal flow in determination of *situs* during development. Furthermore, we propose that radial spoke head proteins can be considered good candidate genes for PCD without *situs inversus*.

A proportion of PCD patients are classed as "radial spoke defect". These patients can have *situs inversus* and they have been shown by EM to have absent radial spokes and this is typically accompanied by absence of the inner dynein arm (Chilvers et al., 2003). In these cases, the central pair microtubules can be affected and take up an eccentric position suggesting that they are destabilised. The beat pattern is "stiff", such that the cilia fail to bend, and the cilia have a greatly reduced beat frequency (Chilvers et al., 2003). For this defect, we can predict that the complete loss of the radial spoke-central pair interactions that usually act to propagate the cilia bend (Section 5.2, Figure 5.12), cause the "stiff" beat pattern, coupled with the loss of inner dynein arm-radial spoke interaction that acts on the velocity of the bend, resulting in the reduced beat frequency. These beat patterns are consistent with the presence of *situs inversus* in such PCD cases with the classic radial spoke defect, since nodal cilia motility would be expected to result in laterality defects. The classic radial spoke defect is a distinct phenotype to that of the Arabic PCD families used in this study who have an intermittent central pair and a circular beat pattern with a normal beat frequency. As described in Section 5.2, these families still have inner dynein arms present in the axoneme, so the ciliary beat frequency is not affected, but

the beat pattern may not be regulated due to disruption in the radial spoke-central pair interactions.

6.1.2 Mapping genes for PCD, a genetically heterogeneous disorder

Positional cloning and mutational screening of candidate genes in PCD laboratories and diagnostic centres has highlighted the high degree of genetic and phenotypic heterogeneity that exists for PCD. Genetic heterogeneity is further proven by the different modes of inheritance of PCD: although the majority of cases are autosomal recessive, there have been some X-linked and autosomal dominant cases documented (Krawczynski and Witt, 2004) (Moore et al., 2005) (Narayan et al., 1994). It has also been proposed that PCD could be inherited as a digenic disorder, whereby two mutations in two different genes cause the PCD phenotype (Blouin et al., 2000). The PCD genes *DNAI1*, *DNAH5*, *DNAH11* and *TXNDC3* have all been associated with genetic heterogeneity: each cause only a proportion of disease in families with similar origins and ultrastructural defects (Pennarun et al., 1999; Bartoloni et al., 2002; Duriez et al., 2007; Olbrich et al., 2002). Mutations in *DNAI1*, *DNAH5* and *TXNDC3* have all been shown to cause missing outer dynein arms and no further sub-grouping of the phenotype according to different gene mutations has yet been possible. Therefore, although grouping of affected individuals based on their ultrastructural defect is likely to reduce heterogeneity, patients classified as having similar ciliary defects may still harbour mutations in different genes. This is a problem of the level of ultrastructural detection that is possible as present. EM results vary and this is likely to be contributed to by the fact that it is mostly based on cross sections and we know that cilia vary along the axoneme (Fliegauf et al., 2005). The inner dynein arm is highly heterogeneous and difficult to visualise, as are radial spokes so these ultrastructural defects may often be undetected. Furthermore, different laboratories perform EM for diagnostic purposes and therefore the results can be inconsistent and of varying quality.

In order to minimise heterogeneity in this study, in Chapter 3, five families (120, 130, 141, 143 and 145 – Figure 3.1) of Pakistani origin with outer and inner dynein arms defects were grouped together for genetic analysis. However, genetic homogeneity

was not guaranteed and indeed, two loci on chromosome 11q23.3-q24.3 and chromosome 17q21.31-q22 were consistent with linkage in four out of five families in both cases. Therefore, despite grouping individuals with the same geographical origin and/or ultrastructural defect, they are not guaranteed to be linked to the same locus. Both loci were statistically significant and indistinguishable by HOMOG3R analysis for likelihood of linkage. Similarly, a microsatellite screen performed on family 146 with three affected individuals (Figure 3.1) from the United Arab Emirates (UAE) with a central pair agenesis defect (Stannard et al., 2004), had revealed two putative loci of equal weight on chromosome 3 and 6 that both achieved LOD scores of 2.5 (Section 4.1.6). Again, these loci were indistinguishable for this family alone.

Therefore, more than one locus of equal significance was found in these family groups, grouped together based on geographical origin, or even as one single family. In the case of the Pakistani families, this happened because the families were split across two loci, whereas in the case of single Arabic family 146, this was due to lack of power for linkage: there were only three affected patients which together could achieve a maximum LOD score of 2.5. In contrast, a SNP screen performed on the large consanguineous family 152 with four affected individuals revealed a clear-cut single locus with a LOD score >3 on chromosome 6, with no other region in the genome achieving a LOD score above 0.5 (Appendix 2). This was due to the greater power afforded by a family with four affected individuals. The locus on chromosome 6 was found to overlap with that previously identified for family 146 and together these two Arabic families achieved a LOD score of 6.7 on chromosome 6, so this allowed prioritisation of this locus rather than the locus on chromosome 3 in family 146. Eventually a common mutation was found within a positional candidate gene in both families 146 and 152. This study therefore demonstrated that despite the presence of consanguinity, larger families with more affected individuals are preferable to multiple families in genetic analysis of such a heterogeneous disorder, in order to be sure of the true linkage status of any loci identified. However, linkage analysis may be undertaken in a family/families that will not achieve a LOD score of 3 and intelligent bioinformatic analysis of ciliary candidates in regions of interest may be identified successfully.

Until recently, microsatellites have been used as the primary markers in linkage analysis. The development of several high through-put genotyping platforms, including that supplied by *Illumina* used in this study, now make it feasible for SNPs to be used in linkage analysis. There has been considerable debate on the advantages and disadvantages of the use of microsatellite versus SNP-based genome-wide linkage screens (Schaid et al., 2004). An advantage of using microsatellites is that they are highly polymorphic, more so than the diallelic SNPs. They produce higher LOD scores compared to SNPs due to increased informativity. However, the advantage of SNPs is that there are more present in the genome so are used in linkage screens on the basis that their greater density will compensate for the smaller amount of information per SNP. Furthermore, a number of SNPs in close proximity in the genome can act as "super" alleles by forming local haplotypes and therefore offer greater linkage content. In Chapter 4 of this study, the region on chromosome 6 that was highlighted following a 600 kb SNP screen and found to be consistent for linkage in Arabic family 152, was previously undetected in a less dense 10 cM microsatellite screen. In this case, therefore, the denser SNP based genome screens were proven to be more valuable than microsatellites in identifying regions of homozygosity, even in a large consanguineous PCD family with four affected individuals.

An increased understanding of the structure and function of human cilia has arisen from model organism studies as well as some significant advances including characterisation of the human cilia proteome. The structure and function of cilia has been investigated extensively using the model organism *Chlamydomonas reinhardtii* due to the high conservation of its flagella proteins with orthologues in the human genome present in human cilia and sperm flagella. In addition, much has been learnt about motile and sensory cilia in mammals using electron microscopy and immunolocalisation studies (Satir and Christensen, 2008). Much more work will be needed to understand the additional complexity of the human system. Using *Chlamydomonas* studies, it is known that dynein regulation in motile cilia and flagella depends upon interactions between the radial spokes and central pair apparatus that govern cilia waveform (Mitchell, 2004). However, the exact nature and function of the components involved is not yet understood for either *Chlamydomonas* flagella or mammalian cilia. In addition, the role of cilia/flagella protein interactions such as these is not understood within embryogenesis and development. Further work is still

needed to enhance the understanding of cilia in both development and disease and offer more candidates in the mapping of genes for ciliopathies.

6.1.3 The future for identification of PCD genes

PCD can be difficult to diagnose since the diagnostic techniques are rather specialised and some of the symptoms such as recurrent respiratory infection and glue ear are common in children (Afzelius, 1998). Moreover, disease severity is highly variable. Early diagnosis leading to early treatment correlates with an improved prognosis and morbidity which allow a better quality of life for patients (Bush et al., 1998; Coren et al., 2002). The identification of disease genes would aid in the diagnosis of PCD in individuals with a history of recurrent respiratory infection. The location of heavy dynein chains DNAH5, DNAH8 and DNAH7 and intermediate chain DNAI1 have been shown by immunofluorescence to be defective in a proportion of PCD patients (Fliegauf et al., 2005; Zhang et al., 2002). Similar proteomic approaches in other PCD patients could potentially identify the absence of particular protein components, provide a firm basis for the sequencing of known PCD genes and even allow for the more accurate grouping of patients to further minimise heterogeneity. This could be aided by development of additional antibodies to axonemal components to complement those already in use, for example by raising radial spoke and central pair antisera.

Since PCD is a recessive disorder, it is most likely to be caused by loss of function of a gene product and therefore gene replacement therapy may in future be appropriate. In addition, the primary defects in PCD affect the structure and/or function of cilia located in the relatively accessible upper and lower respiratory tract. Heterozygous parents of PCD patients are asymptomatic, thus indicating that the protein would only need to be restored to 50 % of wild type levels in order to restore a normal phenotype. The dynamic nature of cilia by intraflagellar transport (IFT) means that the normal gene product could be expected to correctly incorporate into the axoneme and therefore lead to the restoration of functional cilia motility. Our studies that rescued *Chlamydomonas* motility in an RSP9-depleted mutant indicate that this is likely to be the case in RSP9 PCD patients. Progress in the treatment of the respiratory tract disease cystic fibrosis (CF) with gene therapy in humans with CF has been made,

although complications have occurred associated with limitations of effective gene transfer and the requirement of constant re-treatment due to a limited persistence of gene expression (Welsh and Zabner, 1999). However, current advances in gene therapy for respiratory diseases such as CF might pave the way for similar treatments for patients with PCD and as such, identification of genes causing PCD would allow potential gene therapy techniques to be used in the cure for PCD in future. Furthermore, there might also be utility for wider PCD-related phenotypes, including infertility.

The link between left-right body axis determination and motility of embryonic nodal cilia has been documented. It is possible that the identification of genes involved in the structure and function of both respiratory and nodal cilia may allow important insights into the process of laterality determination. It is still not clear if both motile and non-motile cilia co-exist at the node. In particular, the link between the motility of nodal cilia and the initiation of downstream signalling cascades and gene expression requires further understanding. A greater understanding of the pathway of laterality determination would be of particular benefit for patients with PCD and *situs ambiguous*, since this requires clinical interventions unlike *situs inversus*, which is harmless.

Many PCD genes remain to be identified and characterisation of the protein composition of the ciliary axoneme using model organisms such as *Chlamydomonas* has helped with this process. The development of resources such as www.ciliaproteome.org have been hugely powerful tools for PCD gene identification. Further understanding of cilia structure and function will impact on the current understanding of ciliopathies and respiratory research, left-right axis determination, male and female fertility, as well as sensory cilia function within the kidney and photoreceptor cells of the retina.

REFERENCES

Abecasis,G.R., Cherny,S.S., Cookson,W.O., and Cardon,L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. *30*, 97-101.

Afzelius,B.A. (1998). Genetics and pulmonary medicine. 6. Immotile cilia syndrome: past, present, and prospects for the future. Thorax *53*, 894-897.

Afzelius,B.A. (1985). The immotile-cilia syndrome: a microtubule-associated defect. CRC Crit Rev. Biochem. *19*, 63-87.

Afzelius,B.A. (1981). Genetical and ultrastructural aspects of the immotile-cilia syndrome. Am. J. Hum. Genet. *33*, 852-864.

Afzelius,B.A. (1976). A human syndrome caused by immotile cilia. Science *193*, 317-319.

Afzelius,B.A. and Eliasson,R. (1983). Male and female infertility problems in the immotile-cilia syndrome. Eur. J. Respir. Dis. Suppl *127*, 144-147.

Afzelius,B.A., Gargani,G., and Romano,C. (1985). Abnormal length of cilia as a possible cause of defective mucociliary clearance. Eur. J. Respir. Dis. *66*, 173-180.

Antonelli,M., Bravo,E., Quattrucci,S., De,A.M., and Modesti,A. (1982). [The absence of the internal arms of dynein as a cause of the immotile cilia syndrome]. Pediatr. Med. Chir *4*, 639-641.

Arendt,D., Tessmar-Raible,K., Snyman,H., Dorresteijn,A.W., and Wittbrodt,J. (2004). Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science *306*, 869-871.

Asai,D.J. and Wilkes,D.E. (2004). The dynein heavy chain family. J. Eukaryot. Microbiol. *51*, 23-29.

Badano,J.L., Mitsuma,N., Beales,P.L., and Katsanis,N. (2006). The ciliopathies: an emerging class of human genetic disorders. Annu. Rev. Genomics Hum. Genet. *7*, 125-148.

Baker,J.D., Adhikarakunnathu,S., and Kernan,M.J. (2004). Mechanosensory-defective, male-sterile unc mutants identify a novel basal body protein required for ciliogenesis in Drosophila. Development *131*, 3411-3422.

Bartolini,F., Bhamidipati,A., Thomas,S., Schwahn,U., Lewis,S.A., and Cowan,N.J. (2002). Functional overlap between retinitis pigmentosa 2 protein and the tubulin-specific chaperone cofactor C. J. Biol. Chem. *277*, 14629-14634.

Bartolini,F., Tian,G., Piehl,M., Cassimeris,L., Lewis,S.A., and Cowan,N.J. (2005). Identification of a novel tubulin-destabilizing protein related to the chaperone cofactor E. J. Cell Sci. *118*, 1197-1207.

Bartoloni,L., Blouin,J.L., Maiti,A.K., Sainsbury,A., Rossier,C., Gehrig,C., She,J.X., Marron,M.P., Lander,E.S., Meeks,M., Chung,E., Armengot,M., Jorissen,M., Scott,H.S., ozier-Blanchet,C.D., Gardiner,R.M., and Antonarakis,S.E. (2001). Axonemal beta heavy chain dynein DNAH9: cDNA sequence, genomic structure, and investigation of its role in primary ciliary dyskinesia. Genomics *72*, 21-33.

Bartoloni,L., Blouin,J.L., Pan,Y., Gehrig,C., Maiti,A.K., Scamuffa,N., Rossier,C., Jorissen,M., Armengot,M., Meeks,M., Mitchison,H.M., Chung,E.M., ozier-

Blanchet,C.D., Craigen,W.J., and Antonarakis,S.E. (2002). Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. Proc. Natl. Acad. Sci. U. S. A *99*, 10282-10286.

Bellomo,D., Lander,A., Harragan,I., and Brown,N.A. (1996). Cell proliferation in mammalian gastrulation: the ventral node and notochord are relatively quiescent. Dev. Dyn. *205*, 471-485.

Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B., Bohme,U., Hannick,L., Aslett,M.A., Shallom,J., Marcello,L., Hou,L., Wickstead,B., Alsmark,U.C., Arrowsmith,C., Atkin,R.J., Barron,A.J., Bringaud,F., Brooks,K., Carrington,M., Cherevach,I., Chillingworth,T.J., Churcher,C., Clark,L.N., Corton,C.H., Cronin,A., Davies,R.M., Doggett,J., Djikeng,A., Feldblyum,T., Field,M.C., Fraser,A., Goodhead,I., Hance,Z., Harper,D., Harris,B.R., Hauser,H., Hostetler,J., Ivens,A., Jagels,K., Johnson,D., Johnson,J., Jones,K., Kerhornou,A.X., Koo,H., Larke,N., Landfear,S., Larkin,C., Leech,V., Line,A., Lord,A., Macleod,A., Mooney,P.J., Moule,S., Martin,D.M., Morgan,G.W., Mungall,K., Norbertczak,H., Ormond,D., Pai,G., Peacock,C.S., Peterson,J., Quail,M.A., Rabbinowitsch,E., Rajandream,M.A., Reitter,C., Salzberg,S.L., Sanders,M., Schobel,S., Sharp,S., Simmonds,M., Simpson,A.J., Tallon,L., Turner,C.M., Tait,A., Tivey,A.R., Van,A.S., Walker,D., Wanless,D., Wang,S., White,B., White,O., Whitehead,S., Woodward,J., Wortman,J., Adams,M.D., Embley,T.M., Gull,K., Ullu,E., Barry,J.D., Fairlamb,A.H., Opperdoes,F., Barrell,B.G., Donelson,J.E., Hall,N., Fraser,C.M., Melville,S.E., and El-Sayed,N.M. (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416-422.

Bhat,A., Heath,S.C., and Ott,J. (1999). Heterogeneity for multiple disease loci in linkage analysis. Hum. Hered. *49*, 229-231.

Bianchi,E., Savasta,S., Calligaro,A., Beluffi,G., Poggi,P., Tinelli,M., Mevio,E., and Martinetti,M. (1992). HLA haplotype segregation and ultrastructural study in familial immotile-cilia syndrome. Hum. Genet. *89*, 270-274.

Blouin,J.L., Meeks,M., Radhakrishna,U., Sainsbury,A., Gehring,C., Sail,G.D., Bartoloni,L., Dombi,V., O'Rawe,A., Walne,A., Chung,E., Afzelius,B.A., Armengot,M., Jorissen,M., Schidlow,D.V., van,M.L., Walt,H., Gardiner,R.M., Probst,D., Guerne,P.A., ozier-Blanchet,C.D., and Antonarakis,S.E. (2000). Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. Eur. J. Hum. Genet. *8*, 109-118.

Bodini,A., Rugolotto,S., Pradal,U., Zanotto,G., and Peroni,D. (2008). Nasal nitric oxide for early diagnosis of familial primary ciliary dyskinesia. Arch. Dis. Child *93*, 452-453.

Botstein,D., White,R.L., Skolnick,M., and Davis,R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. *32*, 314-331.

Brinkley,W. (1997). Microtubules: a brief historical perspective. J. Struct. Biol. *118*, 84-86.

Broadhead,R., Dawe,H.R., Farr,H., Griffiths,S., Hart,S.R., Portman,N., Shaw,M.K., Ginger,M.L., Gaskell,S.J., McKean,P.G., and Gull,K. (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. Nature *440*, 224-227.

Brokaw,C.J. and Luck,D.J. (1985). Bending patterns of chlamydomonas flagella: III. A radial spoke head deficient mutant and a central pair deficient mutant. Cell Motil. *5*, 195-208.

Brokaw,C.J., Luck,D.J., and Huang,B. (1982). Analysis of the movement of Chlamydomonas flagella:" the function of the radial-spoke system is revealed by comparison of wild-type and mutant flagella. J. Cell Biol. *92*, 722-732.

Bruce,C., Yates,D.H., and Thomas,P.S. (2002). Caffeine decreases exhaled nitric oxide. Thorax *57*, 361-363.

Brueckner,M., D'Eustachio,P., and Horwich,A.L. (1989). Linkage mapping of a mouse gene, iv, that controls left-right asymmetry of the heart and viscera. Proc. Natl. Acad. Sci. U. S. A *86*, 5035-5038.

Bryan,J.H. (1983). The immotile cilia syndrome. Mice versus man. Virchows Arch. A Pathol. Anat. Histopathol. *399*, 265-275.

Budny,B., Chen,W., Omran,H., Fliegauf,M., Tzschach,A., Wisniewska,M., Jensen,L.R., Raynaud,M., Shoichet,S.A., Badura,M., Lenzner,S., Latos-Bielenska,A., and Ropers,H.H. (2006). A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Hum. Genet. *120*, 171-178.

Bush,A. (1998). Congenital heart disease in primary ciliary dyskinesia. Pediatr. Cardiol. *19*, 191.

Bush,A., Chodhari,R., Collins,N., Copeland,F., Hall,P., Harcourt,J., Hariri,M., Hogg,C., Lucas,J., Mitchison,H.M., O'Callaghan,C., and Phillips,G. (2007). Primary ciliary dyskinesia: current state of the art. Archives of Disease in Childhood *92*, 1136-1140.

Bush,A., Cole,P., Hariri,M., Mackay,I., Phillips,G., O'Callaghan,C., Wilson,R., and Warner,J.O. (1998). Primary ciliary dyskinesia: diagnosis and standards of care. Eur. Respir. J. *12*, 982-988.

Carretero Gracia,J.A., Uliarte,R.A., and Martinez-Penuela Virseda,J.M. (2000). [Primary ciliary dyskinesia. A new phenotypic variant]. Arch. Bronconeumol. *36*, 225-227.

Chilvers,M.A. and O'Callaghan,C. (2000). Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparison with the photomultiplier and photodiode methods. Thorax *55*, 314-317.

Chilvers,M.A., Rutman,A., and O'Callaghan,C. (2003). Ciliary beat pattern is associated with specific ultrastructural defects in primary ciliary dyskinesia. J. Allergy Clin. Immunol. *112*, 518-524.

Cole,D.G. (2005). Intraflagellar transport: keeping the motors coordinated. Curr. Biol. *15*, R798-R801.

Cole,D.G. (1999). Kinesin-II, the heteromeric kinesin. Cell Mol. Life Sci. *56*, 217-226.

Collins,J.S. and Schwartz,C.E. (2002). Detecting polymorphisms and mutations in candidate genes. Am. J. Hum. Genet. *71*, 1251-1252.

Coren,M.E., Meeks,M., Morrison,I., Buchdahl,R.M., and Bush,A. (2002). Primary ciliary dyskinesia: age at diagnosis and symptom history. Acta Paediatr. *91*, 667-669.

Cosson,J. (1996). A moving image of flagella: news and views on the mechanisms involved in axonemal beating. Cell Biol. Int. *20*, 83-94.

Davies,J.P. and Grossman,A.R. (1994). Sequences controlling transcription of the Chlamydomonas reinhardtii beta 2-tubulin gene after deflagellation and during the cell cycle. Mol. Cell Biol. *14*, 5165-5174.

Dawe,H.R., Smith,U.M., Cullinane,A.R., Gerrelli,D., Cox,P., Badano,J.L., Blair-Reid,S., Sriram,N., Katsanis,N., ttie-Bitach,T., Afford,S.C., Copp,A.J., Kelly,D.A., Gull,K., and Johnson,C.A. (2007). The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum. Mol. Genet. *16*, 173-186.

de Iongh,R.U. and Rutland,J. (1995). Ciliary defects in healthy subjects, bronchiectasis, and primary ciliary dyskinesia. Am. J. Respir. Crit Care Med. *151*, 1559-1567.

De,S.M., Lobetti,R.G., and Van,W.E. (2004). Primary ciliary dyskinesia in a Staffordshire bull terrier. J. S. Afr. Vet. Assoc. *75*, 150-152.

DiBella,L.M., Gorbatyuk,O., Sakato,M., Wakabayashi,K., Patel-King,R.S., Pazour,G.J., Witman,G.B., and King,S.M. (2005). Differential light chain assembly influences outer arm dynein motor function. Mol. Biol. Cell *16*, 5661-5674.

Duriez,B., Duquesnoy,P., Escudier,E., Bridoux,A.M., Escalier,D., Rayet,I., Marcos,E., Vojtek,A.M., Bercher,J.F., and Amselem,S. (2007). A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. Proc. Natl. Acad. Sci. U. S. A *104*, 3336-3341.

Dutcher,S.K. (1995). Flagellar assembly in two hundred and fifty easy-to-follow steps. Trends Genet. *11*, 398-404.

EBERSOLD,W.T., LEVINE,R.P., LEVINE,E.E., and OLMSTED,M.A. (1962). Linkage maps in Chlamydomonas reinhardi. Genetics *47*, 531-543.

Essner,J.J., Amack,J.D., Nyholm,M.K., Harris,E.B., and Yost,H.J. (2005). Kupffer's vesicle is a ciliated organ of asymmetry in the zebrafish embryo that initiates left-right development of the brain, heart and gut. Development *132*, 1247-1260.

Essner,J.J., Vogan,K.J., Wagner,M.K., Tabin,C.J., Yost,H.J., and Brueckner,M. (2002). Conserved function for embryonic nodal cilia. Nature *418*, 37-38.

Farrall,M. (1993). Homozygosity mapping: familiarity breeds debility. Nat. Genet. *5*, 107-108.

Ferrante,M.I., Giorgio,G., Feather,S.A., Bulfone,A., Wright,V., Ghiani,M., Selicorni,A., Gammaro,L., Scolari,F., Woolf,A.S., Sylvie,O., Bernard,L., Malcolm,S., Winter,R., Ballabio,A., and Franco,B. (2001). Identification of the gene for oral-facial-digital type I syndrome. Am. J. Hum. Genet. *68*, 569-576.

Ferrante,M.I., Zullo,A., Barra,A., Bimonte,S., Messaddeq,N., Studer,M., Dolle,P., and Franco,B. (2006). Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. Nat. Genet. *38*, 112-117.

Fliegauf,M., Benzing,T., and Omran,H. (2007). When cilia go bad: cilia defects and ciliopathies. Nat. Rev. Mol. Cell Biol. *8*, 880-893.

Fliegauf,M., Olbrich,H., Horvath,J., Wildhaber,J.H., Zariwala,M.A., Kennedy,M., Knowles,M.R., and Omran,H. (2005). Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. Am. J. Respir. Crit Care Med. *171*, 1343-1349.

Frayne,J. and Hall,L. (2002). A re-evaluation of sperm protein 17 (Sp17) indicates a regulatory role in an A-kinase anchoring protein complex, rather than a unique role in sperm-zona pellucida binding. Reproduction. *124*, 767-774.

Freshour,J., Yokoyama,R., and Mitchell,D.R. (2007). Chlamydomonas flagellar outer row dynein assembly protein ODA7 interacts with both outer row and I1 inner row dyneins. J. Biol. Chem. *282*, 5404-5412.

Gardner,L.C., O'Toole,E., Perrone,C.A., Giddings,T., and Porter,M.E. (1994). Components of a "dynein regulatory complex" are located at the junction between the radial spokes and the dynein arms in Chlamydomonas flagella. J. Cell Biol. *127*, 1311-1325.

Gebbia,M., Ferrero,G.B., Pilia,G., Bassi,M.T., Aylsworth,A., Penman-Splitt,M., Bird,L.M., Bamforth,J.S., Burn,J., Schlessinger,D., Nelson,D.L., and Casey,B. (1997). X-linked situs abnormalities result from mutations in ZIC3. Nat. Genet. *17*, 305-308.

Geremek,M., Schoenmaker,F., Zietkiewicz,E., Pogorzelski,A., Diehl,S., Wijmenga,C., and Witt,M. (2008). Sequence analysis of 21 genes located in the Kartagener syndrome linkage region on chromosome 15q. Eur. J. Hum. Genet.

Geremek,M. and Witt,M. (2004). Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. J. Appl. Genet. *45*, 347-361.

Geremek,M., Zietkiewicz,E., Diehl,S.R., Alizadeh,B.Z., Wijmenga,C., and Witt,M. (2006). Linkage analysis localises a Kartagener syndrome gene to a 3.5 cM region on chromosome 15q24-25. J. Med. Genet. *43*, e1.

Gomperts,B.N., Gong-Cooper,X., and Hackett,B.P. (2004). Foxj1 regulates basal body anchoring to the cytoskeleton of ciliated pulmonary epithelial cells. J. Cell Sci. *117*, 1329-1337.

Goodenough,U.W. and Heuser,J.E. (1985). Substructure of inner dynein arms, radial spokes, and the central pair/projection complex of cilia and flagella. J. Cell Biol. *100*, 2008-2018.

Grayson,C., Bartolini,F., Chapple,J.P., Willison,K.R., Bhamidipati,A., Lewis,S.A., Luthert,P.J., Hardcastle,A.J., Cowan,N.J., and Cheetham,M.E. (2002). Localization in the human retina of the X-linked retinitis pigmentosa protein RP2, its homologue cofactor C and the RP2 interacting protein Arl3. Hum. Mol. Genet. *11*, 3065-3074.

Grizzi,F., Chiriva-Internati,M., Franceschini,B., Bumm,K., Colombo,P., Ciccarelli,M., Donetti,E., Gagliano,N., Hermonat,P.L., Bright,R.K., Gioia,M., Dioguardi,N., and Kast,W.M. (2004). Sperm protein 17 is expressed in human somatic ciliated epithelia. J. Histochem. Cytochem. *52*, 549-554.

Guichard,C., Harricane,M.C., Lafitte,J.J., Godard,P., Zaegel,M., Tack,V., Lalau,G., and Bouvagnet,P. (2001). Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). Am. J. Hum. Genet. *68*, 1030-1035.

Handel,M.A. and Kennedy,J.R. (1984). Situs inversus in homozygous mice without immotile cilia. J. Hered. *75*, 498.

Hearn,T., Renforth,G.L., Spalluto,C., Hanley,N.A., Piper,K., Brickwood,S., White,C., Connolly,V., Taylor,J.F., Russell-Eggitt,I., Bonneau,D., Walker,M., and Wilson,D.I.

(2002). Mutation of ALMS1, a large gene with a tandem repeat encoding 47 amino acids, causes Alstrom syndrome. Nat. Genet. *31*, 79-83.

Hearn,T., Spalluto,C., Phillips,V.J., Renforth,G.L., Copin,N., Hanley,N.A., and Wilson,D.I. (2005). Subcellular localization of ALMS1 supports involvement of centrosome and basal body dysfunction in the pathogenesis of obesity, insulin resistance, and type 2 diabetes. Diabetes *54*, 1581-1587.

Hildebrandt,F. and Zhou,W. (2007). Nephronophthisis-associated ciliopathies. J. Am. Soc. Nephrol. *18*, 1855-1871.

Hirokawa,N., Tanaka,Y., Okada,Y., and Takeda,S. (2006). Nodal flow and the generation of left-right asymmetry. Cell *125*, 33-45.

Holley,S.A., Julich,D., Rauch,G.J., Geisler,R., and Nusslein-Volhard,C. (2002). her1 and the notch pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. Development *129*, 1175-1183.

Hong,D.H., Pawlyk,B., Sokolov,M., Strissel,K.J., Yang,J., Tulloch,B., Wright,A.F., Arshavsky,V.Y., and Li,T. (2003). RPGR isoforms in photoreceptor connecting cilia and the transitional zone of motile cilia. Invest Ophthalmol. Vis. Sci. *44*, 2413-2421.

Hornef,N., Olbrich,H., Horvath,J., Zariwala,M.A., Fliegauf,M., Loges,N.T., Wildhaber,J., Noone,P.G., Kennedy,M., Antonarakis,S.E., Blouin,J.L., Bartoloni,L., Nublein,T., Ahrens,P., Griese,M., Kuhl,H., Sudbrak,R., Knowles,M.R., Reinhardt,R., and Omran,H. (2006a). DNAH5 Mutations are a Common Cause of Primary Ciliary Dyskinesia with Outer Dynein Arm Defects. Am. J. Respir. Crit Care Med.

Hornef,N., Olbrich,H., Horvath,J., Zariwala,M.A., Fliegauf,M., Loges,N.T., Wildhaber,J., Noone,P.G., Kennedy,M., Antonarakis,S.E., Blouin,J.L., Bartoloni,L., Nusslein,T., Ahrens,P., Griese,M., Kuhl,H., Sudbrak,R., Knowles,M.R., Reinhardt,R., and Omran,H. (2006b). DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. Am. J. Respir. Crit Care Med. *174*, 120-126.

Horvath,J., Fliegauf,M., Olbrich,H., Kispert,A., King,S.M., Mitchison,H., Zariwala,M.A., Knowles,M.R., Sudbrak,R., Fekete,G., Neesen,J., Reinhardt,R., and Omran,H. (2005). Identification and analysis of axonemal dynein light chain 1 in primary ciliary dyskinesia patients. Am. J. Respir. Cell Mol. Biol. *33*, 41-47.

Hou,X., Mrug,M., Yoder,B.K., Lefkowitz,E.J., Kremmidiotis,G., D'Eustachio,P., Beier,D.R., and Guay-Woodford,L.M. (2002). Cystin, a novel cilia-associated protein, is disrupted in the cpk mouse model of polycystic kidney disease. J. Clin. Invest *109*, 533-540.

Huang,B., Piperno,G., Ramanis,Z., and Luck,D.J. (1981). Radial spokes of Chlamydomonas flagella: genetic analysis of assembly and function. J. Cell Biol. *88*, 80-88.

Ibanez-Tallon,I., Gorokhova,S., and Heintz,N. (2002). Loss of function of axonemal dynein Mdnah5 causes primary ciliary dyskinesia and hydrocephalus. Hum. Mol. Genet. *11*, 715-721.

Ibanez-Tallon,I., Heintz,N., and Omran,H. (2003). To beat or not to beat: roles of cilia in development and disease. Hum. Mol. Genet. *12 Spec No 1*, R27-R35.

Ibanez-Tallon,I., Pagenstecher,A., Fliegauf,M., Olbrich,H., Kispert,A., Ketelsen,U.P., North,A., Heintz,N., and Omran,H. (2004). Dysfunction of axonemal dynein heavy chain

Mdnah5 inhibits ependymal flow and reveals a novel mechanism for hydrocephalus formation. Hum. Mol. Genet. *13*, 2133-2141.

Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R., Anupama,A., Apostolou,Z., Attipoe,P., Bason,N., Bauser,C., Beck,A., Beverley,S.M., Bianchettin,G., Borzym,K., Bothe,G., Bruschi,C.V., Collins,M., Cadag,E., Ciarloni,L., Clayton,C., Coulson,R.M., Cronin,A., Cruz,A.K., Davies,R.M., De,G.J., Dobson,D.E., Duesterhoeft,A., Fazelina,G., Fosker,N., Frasch,A.C., Fraser,A., Fuchs,M., Gabel,C., Goble,A., Goffeau,A., Harris,D., Hertz-Fowler,C., Hilbert,H., Horn,D., Huang,Y., Klages,S., Knights,A., Kube,M., Larke,N., Litvin,L., Lord,A., Louie,T., Marra,M., Masuy,D., Matthews,K., Michaeli,S., Mottram,J.C., Muller-Auer,S., Munden,H., Nelson,S., Norbertczak,H., Oliver,K., O'neil,S., Pentony,M., Pohl,T.M., Price,C., Purnelle,B., Quail,M.A., Rabbinowitsch,E., Reinhardt,R., Rieger,M., Rinta,J., Robben,J., Robertson,L., Ruiz,J.C., Rutter,S., Saunders,D., Schafer,M., Schein,J., Schwartz,D.C., Seeger,K., Seyler,A., Sharp,S., Shin,H., Sivam,D., Squares,R., Squares,S., Tosato,V., Vogt,C., Volckaert,G., Wambutt,R., Warren,T., Wedler,H., Woodward,J., Zhou,S., Zimmermann,W., Smith,D.F., Blackwell,J.M., Stuart,K.D., Barrell,B., and Myler,P.J. (2005). The genome of the kinetoplastid parasite, Leishmania major. Science *309*, 436-442.

Jabourian,Z., Lublin,F.D., Adler,A., Gonzales,C., Northrup,B., and Zwillenberg,D. (1986). Hydrocephalus in Kartagener's syndrome. Ear Nose Throat J. *65*, 468-472.

Janke,C., Rogowski,K., Wloga,D., Regnard,C., Kajava,A.V., Strub,J.M., Temurak,N., van,D.J., Boucher,D., van,D.A., Suryavanshi,S., Gaertig,J., and Edde,B. (2005). Tubulin polyglutamylase enzymes are members of the TTL domain protein family. Science *308*, 1758-1762.

Jeganathan,D., Chodhari,R., Meeks,M., Faeroe,O., Smyth,D., Nielsen,K., Amirav,I., Luder,A.S., Bisgaard,H., Gardiner,R.M., Chung,E.M., and Mitchison,H.M. (2004). Loci for primary ciliary dyskinesia map to chromosome 16p12.1-12.2 and 15q13.1-15.1 in Faroe Islands and Israeli Druze genetic isolates. J. Med. Genet. *41*, 233-240.

Karadag,B., James,A.J., Gultekin,E., Wilson,N.M., and Bush,A. (1999). Nasal and lower airway level of nitric oxide in children with primary ciliary dyskinesia. Eur. Respir. J. *13*, 1402-1405.

Katsanis,N. (2006). Ciliary proteins and exencephaly. Nat. Genet. *38*, 135-136.

Katsanis,N., Beales,P.L., Woods,M.O., Lewis,R.A., Green,J.S., Parfrey,P.S., Ansley,S.J., Davidson,W.S., and Lupski,J.R. (2000). Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. Nat. Genet. *26*, 67-70.

Kennedy,M.P., Omran,H., Leigh,M.W., Dell,S., Morgan,L., Molina,P.L., Robinson,B.V., Minnix,S.L., Olbrich,H., Severin,T., Ahrens,P., Lange,L., Morillas,H.N., Noone,P.G., Zariwala,M.A., and Knowles,M.R. (2007). Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. Circulation *115*, 2814-2821.

Kimmel,C.B., Ballard,W.W., Kimmel,S.R., Ullmann,B., and Schilling,T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253-310.

Kindle,K.L. (1990). High-frequency nuclear transformation of Chlamydomonas reinhardtii. Proc. Natl. Acad. Sci. U. S. A *87*, 1228-1232.

Kispert,A., Petry,M., Olbrich,H., Volz,A., Ketelsen,U.P., Horvath,J., Melkaoui,R., Omran,H., Zariwala,M., Noone,P.G., and Knowles,M. (2003). Genotype-phenotype correlations in PCD patients carrying DNAH5 mutations. Thorax *58*, 552-554.

Kong,A. and Cox,N.J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. Am. J. Hum. Genet. *61*, 1179-1188.

Kong,A., Gudbjartsson,D.F., Sainz,J., Jonsdottir,G.M., Gudjonsson,S.A., Richardsson,B., Sigurdardottir,S., Barnard,J., Hallbeck,B., Masson,G., Shlien,A., Palsson,S.T., Frigge,M.L., Thorgeirsson,T.E., Gulcher,J.R., and Stefansson,K. (2002). A high-resolution recombination map of the human genome. Nat. Genet. *31*, 241-247.

Kong,M., Richardson,R.T., Widgren,E.E., and O'Rand,M.G. (1995). Sequence and localization of the mouse sperm autoantigenic protein, Sp17. Biol. Reprod. *53*, 579-590.

Kortazar,D., Fanarraga,M.L., Carranza,G., Bellido,J., Villegas,J.C., Avila,J., and Zabala,J.C. (2007). Role of cofactors B (TBCB) and E (TBCE) in tubulin heterodimer dissociation. Exp. Cell Res. *313*, 425-436.

Kosaki,K., Ikeda,K., Miyakoshi,K., Ueno,M., Kosaki,R., Takahashi,D., Tanaka,M., Torikata,C., Yoshimura,Y., and Takahashi,T. (2004). Absent inner dynein arms in a fetus with familial hydrocephalus-situs abnormality. Am. J. Med. Genet. A *129*, 308-311.

Kozminski,K.G., Johnson,K.A., Forscher,P., and Rosenbaum,J.L. (1993). A motility in the eukaryotic flagellum unrelated to flagellar beating. Proc. Natl. Acad. Sci. U. S. A *90*, 5519-5523.

Kramer-Zucker,A.G., Olale,F., Haycraft,C.J., Yoder,B.K., Schier,A.F., and Drummond,I.A. (2005). Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis. Development *132*, 1907-1921.

Krawczynski,M.R. and Witt,M. (2004). PCD and RP: X-linked inheritance of both disorders? Pediatr. Pulmonol. *38*, 88-89.

Lander,E.S. and Botstein,D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science *236*, 1567-1570.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczky,J., LeVine,R., McEwan,P., McKernan,K., Meldrim,J., Mesirov,J.P., Miranda,C., Morris,W., Naylor,J., Raymond,C., Rosetti,M., Santos,R., Sheridan,A., Sougnez,C., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Wyman,D., Rogers,J., Sulston,J., Ainscough,R., Beck,S., Bentley,D., Burton,J., Clee,C., Carter,N., Coulson,A., Deadman,R., Deloukas,P., Dunham,A., Dunham,I., Durbin,R., French,L., Grafham,D., Gregory,S., Hubbard,T., Humphray,S., Hunt,A., Jones,M., Lloyd,C., McMurray,A., Matthews,L., Mercer,S., Milne,S., Mullikin,J.C., Mungall,A., Plumb,R., Ross,M., Shownkeen,R., Sims,S., Waterston,R.H., Wilson,R.K., Hillier,L.W., McPherson,J.D., Marra,M.A., Mardis,E.R., Fulton,L.A., Chinwalla,A.T., Pepin,K.H., Gish,W.R., Chissoe,S.L., Wendl,M.C., Delehaunty,K.D., Miner,T.L., Delehaunty,A., Kramer,J.B., Cook,L.L., Fulton,R.S., Johnson,D.L., Minx,P.J., Clifton,S.W., Hawkins,T., Branscomb,E., Predki,P., Richardson,P., Wenning,S., Slezak,T., Doggett,N., Cheng,J.F., Olsen,A., Lucas,S., Elkin,C., Uberbacher,E., Frazier,M., Gibbs,R.A., Muzny,D.M., Scherer,S.E., Bouck,J.B., Sodergren,E.J., Worley,K.C., Rives,C.M., Gorrell,J.H., Metzker,M.L., Naylor,S.L., Kucherlapati,R.S., Nelson,D.L., Weinstock,G.M., Sakaki,Y., Fujiyama,A., Hattori,M., Yada,T., Toyoda,A., Itoh,T., Kawagoe,C., Watanabe,H., Totoki,Y., Taylor,T., Weissenbach,J., Heilig,R., Saurin,W.,

Artiguenave,F., Brottier,P., Bruls,T., Pelletier,E., Robert,C., Wincker,P., Smith,D.R., Doucette-Stamm,L., Rubenfield,M., Weinstock,K., Lee,H.M., Dubois,J., Rosenthal,A., Platzer,M., Nyakatura,G., Taudien,S., Rump,A., Yang,H., Yu,J., Wang,J., Huang,G., Gu,J., Hood,L., Rowen,L., Madan,A., Qin,S., Davis,R.W., Federspiel,N.A., Abola,A.P., Proctor,M.J., Myers,R.M., Schmutz,J., Dickson,M., Grimwood,J., Cox,D.R., Olson,M.V., Kaul,R., Raymond,C., Shimizu,N., Kawasaki,K., Minoshima,S., Evans,G.A., Athanasiou,M., Schultz,R., Roe,B.A., Chen,F., Pan,H., Ramser,J., Lehrach,H., Reinhardt,R., McCombie,W.R., de la,B.M., Dedhia,N., Blocker,H., Hornischer,K., Nordsiek,G., Agarwala,R., Aravind,L., Bailey,J.A., Bateman,A., Batzoglou,S., Birney,E., Bork,P., Brown,D.G., Burge,C.B., Cerutti,L., Chen,H.C., Church,D., Clamp,M., Copley,R.R., Doerks,T., Eddy,S.R., Eichler,E.E., Furey,T.S., Galagan,J., Gilbert,J.G., Harmon,C., Hayashizaki,Y., Haussler,D., Hermjakob,H., Hokamp,K., Jang,W., Johnson,L.S., Jones,T.A., Kasif,S., Kaspryzk,A., Kennedy,S., Kent,W.J., Kitts,P., Koonin,E.V., Korf,I., Kulp,D., Lancet,D., Lowe,T.M., McLysaght,A., Mikkelsen,T., Moran,J.V., Mulder,N., Pollara,V.J., Ponting,C.P., Schuler,G., Schultz,J., Slater,G., Smit,A.F., Stupka,E., Szustakowski,J., Thierry-Mieg,D., Thierry-Mieg,J., Wagner,L., Wallis,J., Wheeler,R., Williams,A., Wolf,Y.I., Wolfe,K.H., Yang,S.P., Yeh,R.F., Collins,F., Guyer,M.S., Peterson,J., Felsenfeld,A., Wetterstrand,K.A., Patrinos,A., Morgan,M.J., de,J.P., Catanese,J.J., Osoegawa,K., Shizuya,H., Choi,S., and Chen,Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lea,I.A., Widgren,E.E., and O'Rand,M.G. (2004). Association of sperm protein 17 with A-kinase anchoring protein 3 in flagella. Reprod. Biol. Endocrinol. *2*, 57.

Lechtreck,K.F. and Witman,G.B. (2007). Chlamydomonas reinhardtii hydin is a central pair protein required for flagellar motility. J. Cell Biol. *176*, 473-482.

Lee,L., Campagna,D.R., Pinkus,J.L., Mulhern,H., Wyatt,T.A., Sisson,J.H., Pavlik,J.A., Pinkus,G.S., and Fleming,M.D. (2008). Primary ciliary dyskinesia in mice lacking the novel ciliary protein Pcdp1. Mol. Cell Biol. *28*, 949-957.

Lefebvre,P.A. and Silflow,C.D. (1999). Chlamydomonas: the cell and its genomes. Genetics *151*, 9-14.

Leigh,M.W. (2003). Primary ciliary dyskinesia. Semin. Respir. Crit Care Med. *24*, 653-662.

Lemire,M. (2006). SUP: an extension to SLINK to allow a larger number of marker loci to be simulated in pedigrees conditional on trait values. BMC Genet. *7*, 40.

Li,J.B., Gerdes,J.M., Haycraft,C.J., Fan,Y., Teslovich,T.M., May-Simera,H., Li,H., Blacque,O.E., Li,L., Leitch,C.C., Lewis,R.A., Green,J.S., Parfrey,P.S., Leroux,M.R., Davidson,W.S., Beales,P.L., Guay-Woodford,L.M., Yoder,B.K., Stormo,G.D., Katsanis,N., and Dutcher,S.K. (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell *117*, 541-552.

Lieschke,G.J. and Currie,P.D. (2007). Animal models of human disease: zebrafish swim into view. Nat. Rev. Genet. *8*, 353-367.

Lo,K.W., Kan,H.M., and Pfister,K.K. (2006). Identification of a novel region of the cytoplasmic dynein intermediate chain important for dimerization in the absence of the light chains. J. Biol. Chem.

Loukides,S., Kharitonov,S., Wodehouse,T., Cole,P.J., and Barnes,P.J. (1998). Effect of arginine on mucociliary function in primary ciliary dyskinesia. Lancet *352*, 371-372.

Maiti,A.K., Bartoloni,L., Mitchison,H.M., Meeks,M., Chung,E., Spiden,S., Gehrig,C., Rossier,C., Lozier-Blanchet,C.D., Blouin,J., Gardiner,R.M., and Antonarakis,S.E. (2000). No deleterious mutations in the FOXJ1 (alias HFH-4) gene in patients with primary ciliary dyskinesia (PCD). Cytogenet. Cell Genet. *90*, 119-122.

Mandelkow,E. and Mandelkow,E.M. (1995). Microtubules and microtubule-associated proteins. Curr. Opin. Cell Biol. *7*, 72-81.

Marshall,W.F. (2008). The cell biological basis of ciliary disease. J. Cell Biol. *180*, 17-21.

Marszalek,J.R., Liu,X., Roberts,E.A., Chui,D., Marth,J.D., Williams,D.S., and Goldstein,L.S. (2000). Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. Cell *102*, 175-187.

McAuley,J.R. and Anand,V.K. (1998). Clinical significance of compound cilia. Otolaryngol. Head Neck Surg. *118*, 685-687.

McClintock,T.S., Glasser,C.E., Bose,S.C., and Bergman,D.A. (2008). Tissue expression patterns identify mouse cilia genes. Physiol Genomics *32*, 198-206.

McGrath,J. and Brueckner,M. (2003). Cilia are at the heart of vertebrate left-right asymmetry. Curr. Opin. Genet. Dev. *13*, 385-392.

McMahon,A.P., Ingham,P.W., and Tabin,C.J. (2003). Developmental roles and clinical significance of hedgehog signaling. Curr. Top. Dev. Biol. *53*, 1-114.

McManus,I.C., Mitchison,H.M., Chung,E.M., Stubbings,G.F., and Martin,N. (2003). Primary ciliary dyskinesia (Siewert's/Kartagener's syndrome): respiratory symptoms and psycho-social impact. BMC Pulm. Med. *3*, 4.

Meeks,M. and Bush,A. (2000). Primary ciliary dyskinesia (PCD). Pediatr. Pulmonol. *29*, 307-316.

Meeks,M., Walne,A., Spiden,S., Simpson,H., Mussaffi-Georgy,H., Hamam,H.D., Fehaid,E.L., Cheehab,M., Al-Dabbagh,M., Polak-Charcon,S., Blau,H., O'Rawe,A., Mitchison,H.M., Gardiner,R.M., and Chung,E. (2000). A locus for primary ciliary dyskinesia maps to chromosome 19q. J. Med. Genet. *37*, 241-244.

Merchant,S.S., Prochnik,S.E., Vallon,O., Harris,E.H., Karpowicz,S.J., Witman,G.B., Terry,A., Salamov,A., Fritz-Laylin,L.K., Marechal-Drouard,L., Marshall,W.F., Qu,L.H., Nelson,D.R., Sanderfoot,A.A., Spalding,M.H., Kapitonov,V.V., Ren,Q., Ferris,P., Lindquist,E., Shapiro,H., Lucas,S.M., Grimwood,J., Schmutz,J., Cardol,P., Cerutti,H., Chanfreau,G., Chen,C.L., Cognat,V., Croft,M.T., Dent,R., Dutcher,S., Fernandez,E., Fukuzawa,H., Gonzalez-Ballester,D., Gonzalez-Halphen,D., Hallmann,A., Hanikenne,M., Hippler,M., Inwood,W., Jabbari,K., Kalanon,M., Kuras,R., Lefebvre,P.A., Lemaire,S.D., Lobanov,A.V., Lohr,M., Manuell,A., Meier,I., Mets,L., Mittag,M., Mittelmeier,T., Moroney,J.V., Moseley,J., Napoli,C., Nedelcu,A.M., Niyogi,K., Novoselov,S.V., Paulsen,I.T., Pazour,G., Purton,S., Ral,J.P., Riano-Pachon,D.M., Riekhof,W., Rymarquis,L., Schroda,M., Stern,D., Umen,J., Willows,R., Wilson,N., Zimmer,S.L., Allmer,J., Balk,J., Bisova,K., Chen,C.J., Elias,M., Gendler,K., Hauser,C., Lamb,M.R., Ledford,H., Long,J.C., Minagawa,J., Page,M.D., Pan,J., Pootakham,W., Roje,S., Rose,A., Stahlberg,E., Terauchi,A.M., Yang,P., Ball,S., Bowler,C., Dieckmann,C.L., Gladyshev,V.N., Green,P., Jorgensen,R., Mayfield,S., Mueller-Roeber,B., Rajamani,S., Sayre,R.T., Brokstein,P., Dubchak,I., Goodstein,D., Hornick,L., Huang,Y.W., Jhaveri,J., Luo,Y., Martinez,D., Ngau,W.C., Otillar,B., Poliakov,A., Porter,A., Szajkowski,L., Werner,G., Zhou,K., Grigoriev,I.V.,

Rokhsar,D.S., and Grossman,A.R. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245-250.

Miano,M.G., Jacobson,S.G., Carothers,A., Hanson,I., Teague,P., Lovell,J., Cideciyan,A.V., Haider,N., Stone,E.M., Sheffield,V.C., and Wright,A.F. (2000). Pitfalls in homozygosity mapping. Am. J. Hum. Genet. *67*, 1348-1351.

Milisav,I. (1998). Dynein and dynein-related genes. Cell Motil. Cytoskeleton *39*, 261-272.

Mitchell,D.R. (2003). Reconstruction of the projection periodicity and surface architecture of the flagellar central pair complex. Cell Motil. Cytoskeleton *55*, 188-199.

Mitchell,D.R. (2004). Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. Biol. Cell *96*, 691-696.

Mochizuki,T., Wu,G., Hayashi,T., Xenophontos,S.L., Veldhuisen,B., Saris,J.J., Reynolds,D.M., Cai,Y., Gabow,P.A., Pierides,A., Kimberling,W.J., Breuning,M.H., Deltas,C.C., Peters,D.J., and Somlo,S. (1996). PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science *272*, 1339-1342.

Moore,A., Escudier,E., Roger,G., Tamalet,A., Pelosse,B., Marlin,S., Clement,A., Geremek,M., Delaisi,B., Bridoux,A.M., Coste,A., Witt,M., Duriez,B., and Amselem,S. (2005). RPGR is mutated in patients with a complex X-linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. J. Med. Genet.

Morton,N.E. (1955). Sequential tests for the detection of linkage. Am. J. Hum. Genet. *7*, 277-318.

Munro,N.C., Currie,D.C., Lindsay,K.S., Ryder,T.A., Rutman,A., Dewar,A., Greenstone,M.A., Hendry,W.F., and Cole,P.J. (1994). Fertility in men with primary ciliary dyskinesia presenting with respiratory infection. Thorax *49*, 684-687.

Narayan,D., Krishnan,S.N., Upender,M., Ravikumar,T.S., Mahoney,M.J., Dolan,T.F., Jr., Teebi,A.S., and Haddad,G.G. (1994). Unusual inheritance of primary ciliary dyskinesia (Kartagener's syndrome). J. Med. Genet. *31*, 493-496.

Nauli,S.M., Alenghat,F.J., Luo,Y., Williams,E., Vassilev,P., Li,X., Elia,A.E., Lu,W., Brown,E.M., Quinn,S.J., Ingber,D.E., and Zhou,J. (2003). Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. Nat. Genet. *33*, 129-137.

Neesen,J., Drenckhahn,J.D., Tiede,S., Burfeind,P., Grzmil,M., Konietzko,J., Dixkens,C., Kreutzberger,J., Laccone,F., and Omran,H. (2002). Identification of the human ortholog of the t-complex-encoded protein TCTE3 and evaluation as a candidate gene for primary ciliary dyskinesia. Cytogenet. Genome Res. *98*, 38-44.

Nicastro,D., Schwartz,C., Pierson,J., Gaudette,R., Porter,M.E., and McIntosh,J.R. (2006). The molecular architecture of axonemes revealed by cryoelectron tomography. Science *313*, 944-948.

Nievergelt,C.M., Smith,D.W., Kohlenberg,J.B., and Schork,N.J. (2004). Large-scale integration of human genetic and physical maps. Genome Res. *14*, 1199-1205.

Nonaka,S., Tanaka,Y., Okada,Y., Takeda,S., Harada,A., Kanai,Y., Kido,M., and Hirokawa,N. (1998). Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell *95*, 829-837.

Noone,P.G., Leigh,M.W., Sannuti,A., Minnix,S.L., Carson,J.L., Hazucha,M., Zariwala,M.A., and Knowles,M.R. (2004). Primary ciliary dyskinesia: diagnostic and phenotypic features. Am. J. Respir. Crit Care Med. *169*, 459-467.

Okada,Y., Takeda,S., Tanaka,Y., Belmonte,J.C., and Hirokawa,N. (2005). Mechanism of nodal flow: a conserved symmetry breaking event in left-right axis determination. Cell *121*, 633-644.

Olbrich,H., Haffner,K., Kispert,A., Volkel,A., Volz,A., Sasmaz,G., Reinhardt,R., Hennig,S., Lehrach,H., Konietzko,N., Zariwala,M., Noone,P.G., Knowles,M., Mitchison,H.M., Meeks,M., Chung,E.M., Hildebrandt,F., Sudbrak,R., and Omran,H. (2002). Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nat. Genet. *30*, 143-144.

Omran,H., Haffner,K., Volkel,A., Kuehr,J., Ketelsen,U.P., Ross,U.H., Konietzko,N., Wienker,T., Brandis,M., and Hildebrandt,F. (2000). Homozygosity mapping of a gene locus for primary ciliary dyskinesia on chromosome 5p and identification of the heavy dynein chain DNAH5 as a candidate gene. Am. J. Respir. Cell Mol. Biol. *23*, 696-702.

Ostrowski,L.E., Blackburn,K., Radde,K.M., Moyer,M.B., Schlatzer,D.M., Moseley,A., and Boucher,R.C. (2002). A proteomic analysis of human cilia: identification of novel components. Mol. Cell Proteomics *1*, 451-465.

Ott,J. (1999). Methods of analysis and resources available for genetic trait mapping. J. Hered. *90*, 68-70.

Otto,E.A., Loeys,B., Khanna,H., Hellemans,J., Sudbrak,R., Fan,S., Muerb,U., O'Toole,J.F., Helou,J., Attanasio,M., Utsch,B., Sayer,J.A., Lillo,C., Jimeno,D., Coucke,P., De,P.A., Reinhardt,R., Klages,S., Tsuda,M., Kawakami,I., Kusakabe,T., Omran,H., Imm,A., Tippens,M., Raymond,P.A., Hill,J., Beales,P., He,S., Kispert,A., Margolis,B., Williams,D.S., Swaroop,A., and Hildebrandt,F. (2005). Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin. Nat. Genet. *37*, 282-288.

Palmblad,J., Mossberg,B., and Afzelius,B.A. (1984). Ultrastructural, cellular, and clinical features of the immotile-cilia syndrome. Annu. Rev. Med. *35*, 481-492.

Pan,J., Wang,Q., and Snell,W.J. (2005). Cilium-generated signaling and cilia-related disorders. Lab Invest *85*, 452-463.

Park,T.J., Haigo,S.L., and Wallingford,J.B. (2006). Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. Nat. Genet. *38*, 303-311.

Pazour,G.J. (2004). Intraflagellar transport and cilia-dependent renal disease: the ciliary hypothesis of polycystic kidney disease. J. Am. Soc. Nephrol. *15*, 2528-2536.

Pazour,G.J., Agrin,N., Leszyk,J., and Witman,G.B. (2005). Proteomic analysis of a eukaryotic cilium. J. Cell Biol. *170*, 103-113.

Pazour,G.J., Agrin,N., Walker,B.L., and Witman,G.B. (2006). Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. J. Med. Genet. *43*, 62-73.

Pedersen,L.B., Geimer,S., and Rosenbaum,J.L. (2006). Dissecting the molecular mechanisms of intraflagellar transport in chlamydomonas. Curr. Biol. *16*, 450-459.

Pennarun,G., Bridoux,A.M., Escudier,E., stot-Le,M.F., Cacheux,V., Amselem,S., and Duriez,B. (2002). Isolation and expression of the human hPF20 gene orthologous to Chlamydomonas PF20: evaluation as a candidate for axonemal defects of respiratory cilia and sperm flagella. Am. J. Respir. Cell Mol. Biol. *26*, 362-370.

Pennarun,G., Chapelin,C., Escudier,E., Bridoux,A.M., Dastot,F., Cacheux,V., Goossens,M., Amselem,S., and Duriez,B. (2000). The human dynein intermediate chain 2 gene (DNAI2): cloning, mapping, expression pattern, and evaluation as a candidate for primary ciliary dyskinesia. Hum. Genet. *107*, 642-649.

Pennarun,G., Escudier,E., Chapelin,C., Bridoux,A.M., Cacheux,V., Roger,G., Clement,A., Goossens,M., Amselem,S., and Duriez,B. (1999). Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. Am. J. Hum. Genet. *65*, 1508-1519.

Piacentini,G.L., Bodini,A., Peroni,D., Rigotti,E., Pigozzi,R., Pradal,U., and Boner,A.L. (2008). Nasal nitric oxide for early diagnosis of primary ciliary dyskinesia: practical issues in children. Respir. Med. *102*, 541-547.

Picco,P., Leveratto,L., Cama,A., Vigliarolo,M.A., Levato,G.L., Gattorno,M., Zammarchi,E., and Donati,M.A. (1993). Immotile cilia syndrome associated with hydrocephalus and precocious puberty: a case report. Eur. J. Pediatr. Surg. *3 Suppl 1*, 20-21.

Piperno,G., Ramanis,Z., Smith,E.F., and Sale,W.S. (1990). Three distinct inner dynein arms in Chlamydomonas flagella: molecular composition and location in the axoneme. J. Cell Biol. *110*, 379-389.

Pirner,M.A. and Linck,R.W. (1994). Tektins are heterodimeric polymers in flagellar microtubules with axial periodicities matching the tubulin lattice. J. Biol. Chem. *269*, 31800-31806.

Porter,M.E., Bower,R., Knott,J.A., Byrd,P., and Dentler,W. (1999). Cytoplasmic dynein heavy chain 1b is required for flagellar assembly in Chlamydomonas. Mol. Biol. Cell *10*, 693-712.

Porter,M.E. and Johnson,K.A. (1989). Dynein structure and function. Annu. Rev. Cell Biol. *5*, 119-151.

Puffenberger,E.G., Hu-Lince,D., Parod,J.M., Craig,D.W., Dobrin,S.E., Conway,A.R., Donarum,E.A., Strauss,K.A., Dunckley,T., Cardenas,J.F., Melmed,K.R., Wright,C.A., Liang,W., Stafford,P., Flynn,C.R., Morton,D.H., and Stephan,D.A. (2004). Mapping of sudden infant death with dysgenesis of the testes syndrome (SIDDT) by a SNP genome scan and identification of TSPYL loss of function. Proc. Natl. Acad. Sci. U. S. A *101*, 11689-11694.

Roepman,R. and Wolfrum,U. (2007). Protein networks and complexes in photoreceptor cilia. Subcell. Biochem. *43*, 209-235.

Roperto,F., Galati,P., and Rossacco,P. (1993). Immotile cilia syndrome in pigs. A model for human disease. Am. J. Pathol. *143*, 643-647.

Roperto,F., Rossacco,P., Tartaro,A., and Galati,P. (1994). Abnormal length of respiratory cilia in a pig. An ultrastructural study. J. Submicrosc. Cytol. Pathol. *26*, 75-78.

Roth,Y., Baum,G.L., and Tadmor,R. (1988). Brain dysfunction in primary ciliary dyskinesia? Acta Neurol. Scand. *78*, 353-357.

Rozen,S. and Skaletsky,H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. *132*, 365-386.

Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A., and Arnheim,N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science *230*, 1350-1354.

Salisbury,J.L. (2004). Primary cilia: putting sensors together. Curr. Biol. *14*, R765-R767.

Sanderson,M.J. and Sleigh,M.A. (1981). Ciliary activity of cultured rabbit tracheal epithelium: beat pattern and metachrony. J. Cell Sci. *47*, 331-347.

Sapiro,R., Kostetskii,I., Olds-Clarke,P., Gerton,G.L., Radice,G.L., and Strauss III,J.F. (2002). Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. Mol. Cell Biol. *22*, 6298-6305.

Satir,P. and Christensen,S.T. (2008). Structure and function of mammalian cilia. Histochem. Cell Biol.

Saunier,S., Salomon,R., and Antignac,C. (2005). Nephronophthisis. Curr. Opin. Genet. Dev. *15*, 324-331.

Schaid,D.J., Guenther,J.C., Christensen,G.B., Hebbring,S., Rosenow,C., Hilker,C.A., McDonnell,S.K., Cunningham,J.M., Slager,S.L., Blute,M.L., and Thibodeau,S.N. (2004). Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility Loci. Am. J. Hum. Genet. *75*, 948-965.

Schipor,I., Palmer,J.N., Cohen,A.S., and Cohen,N.A. (2006). Quantification of ciliary beat frequency in sinonasal epithelial cells using differential interference contrast microscopy and high-speed digital video imaging. Am. J. Rhinol. *20*, 124-127.

Schneeberger,E.E., McCormack,J., Issenberg,H.J., Schuster,S.R., and Gerald,P.S. (1980). Heterogeneity of ciliary morphology in the immotile-cilia syndrome in man. J. Ultrastruct. Res. *73*, 34-43.

Scholey,J.M. (2008). Intraflagellar transport motors in cilia: moving along the cell's antenna. J. Cell Biol. *180*, 23-29.

Schwabe,G.C., Hoffmann,K., Loges,N.T., Birker,D., Rossier,C., de Santi,M.M., Olbrich,H., Fliegauf,M., Failly,M., Liebers,U., Collura,M., Gaedicke,G., Mundlos,S., Wahn,U., Blouin,J.L., Niggemann,B., Omran,H., Antonarakis,S.E., and Bartoloni,L. (2008). Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. Hum. Mutat. *29*, 289-298.

Shen,R., Fan,J.B., Campbell,D., Chang,W., Chen,J., Doucet,D., Yeakley,J., Bibikova,M., Wickham,G.E., McBride,C., Steemers,F., Garcia,F., Kermani,B.G., Gunderson,K., and Oliphant,A. (2005). High-throughput SNP genotyping on universal bead arrays. Mutat. Res. *573*, 70-82.

Sleigh,M.A. (1981). Primary ciliary dyskinesia. Lancet *2*, 476.

Sloboda,R.D. (2005). Intraflagellar transport and the flagellar tip complex. J. Cell Biochem. *94*, 266-272.

Sloboda,R.D. (2002). A healthy understanding of intraflagellar transport. Cell Motil. Cytoskeleton *52*, 1-8.

Smith,E.F. (2002). Regulation of flagellar dynein by the axonemal central apparatus. Cell Motil. Cytoskeleton *52*, 33-42.

Smith,E.F. and Lefebvre,P.A. (1997). The role of central apparatus components in flagellar motility and microtubule assembly. Cell Motil. Cytoskeleton *38*, 1-8.

Smith,E.F. and Sale,W.S. (1992). Structural and functional reconstitution of inner dynein arms in Chlamydomonas flagellar axonemes. J. Cell Biol. *117*, 573-581.

Smith,E.F. and Yang,P. (2004). The radial spokes and central apparatus: mechano-chemical transducers that regulate flagellar motility. Cell Motil. Cytoskeleton *57*, 8-17.

Smith,U.M., Consugar,M., Tee,L.J., McKee,B.M., Maina,E.N., Whelan,S., Morgan,N.V., Goranson,E., Gissen,P., Lilliquist,S., Aligianis,I.A., Ward,C.J., Pasha,S., Punyashthiti,R., Malik,S.S., Batman,P.A., Bennett,C.P., Woods,C.G., McKeown,C., Bucourt,M., Miller,C.A., Cox,P., Algazali,L., Trembath,R.C., Torres,V.E., ttie-Bitach,T., Kelly,D.A., Maher,E.R., Gattone,V.H., Harris,P.C., and Johnson,C.A. (2006). The transmembrane protein meckelin (MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat. Nat. Genet. *38*, 191-196.

Stannard,W., Rutman,A., Wallis,C., and O'Callaghan,C. (2004). Central microtubular agenesis causing primary ciliary dyskinesia. Am. J. Respir. Crit Care Med. *169*, 634-637.

Stolc,V., Samanta,M.P., Tongprasit,W., and Marshall,W.F. (2005). Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. Proc. Natl. Acad. Sci. U. S. A *102*, 3703-3707.

Strauch,K., Fimmers,R., Kurz,T., Deichmann,K.A., Wienker,T.F., and Baur,M.P. (2000). Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. Am. J. Hum. Genet. *66*, 1945-1957.

Sturgess,J.M., Chao,J., Wong,J., Aspin,N., and Turner,J.A. (1979). Cilia with defective radial spokes: a cause of human respiratory disease. N. Engl. J. Med. *300*, 53-56.

Sullivan-Brown,J., Schottenfeld,J., Okabe,N., Hostetter,C.L., Serluca,F.C., Thiberge,S.Y., and Burdine,R.D. (2008). Zebrafish mutations affecting cilia motility share similar cystic phenotypes and suggest a mechanism of cyst formation that differs from pkd2 morphants. Dev. Biol. *314*, 261-275.

Tabin,C.J. and Vogan,K.J. (2003). A two-cilia model for vertebrate left-right axis specification. Genes Dev. *17*, 1-6.

Takada,S., Wilkerson,C.G., Wakabayashi,K., Kamiya,R., and Witman,G.B. (2002). The outer dynein arm-docking complex: composition and characterization of a subunit (oda1) necessary for outer arm assembly. Mol. Biol. Cell *13*, 1015-1029.

Takeda,S., Yonekawa,Y., Tanaka,Y., Okada,Y., Nonaka,S., and Hirokawa,N. (1999). Left-right asymmetry and kinesin superfamily protein KIF3A: new insights in determination of laterality and mesoderm induction by kif3A-/- mice analysis. J. Cell Biol. *145*, 825-836.

Tamakoshi,T., Itakura,T., Chandra,A., Uezato,T., Yang,Z., Xue,X.D., Wang,B., Hackett,B.P., Yokoyama,T., and Miura,N. (2006). Roles of the Foxj1 and Inv genes in the left-right determination of internal organs in mice. Biochem. Biophys. Res. Commun. *339*, 932-938.

Tanaka,Y., Okada,Y., and Hirokawa,N. (2005). FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. Nature *435*, 172-177.

Teebi,A.S. and El-Shanti,H.I. (2006). Consanguinity: implications for practice, research, and policy. Lancet *367*, 970-971.

Thiele,H. and Nurnberg,P. (2005). HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics. *21*, 1730-1732.

Torban,E., Kor,C., and Gros,P. (2004a). Van Gogh-like2 (Strabismus) and its role in planar cell polarity and convergent extension in vertebrates. Trends Genet. *20*, 570-577.

Torban,E., Wang,H.J., Groulx,N., and Gros,P. (2004b). Independent mutations in mouse Vangl2 that cause neural tube defects in looptail mice impair interaction with members of the Dishevelled family. J. Biol. Chem. *279*, 52703-52713.

Torikata,C., Kijimoto,C., and Koto,M. (1991). Ultrastructure of respiratory cilia of WIC-Hyd male rats. An animal model for human immotile cilia syndrome. Am. J. Pathol. *138*, 341-347.

van's Gravesande and Omran,H. (2005). Primary ciliary dyskinesia: Clinical presentation, diagnosis and genetics. Annals of Medicine *37*, 439-449.

van,R.E., Giles,R.H., Voest,E.E., van,R.C., Schulte-Merker,S., and van Eeden,F.J. (2008). LRRC50, a Conserved Ciliary Protein Implicated in Polycystic Kidney Disease. J. Am. Soc. Nephrol.

Vent,J., Wyatt,T.A., Smith,D.D., Banerjee,A., Luduena,R.F., Sisson,J.H., and Hallworth,R. (2005). Direct involvement of the isotype-specific C-terminus of beta tubulin in ciliary beating. J. Cell Sci. *118*, 4333-4341.

Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., Chen,L., Skupski,M., Subramanian,G., Thomas,P.D., Zhang,J., Gabor Miklos,G.L., Nelson,C., Broder,S., Clark,A.G., Nadeau,J., McKusick,V.A., Zinder,N., Levine,A.J., Roberts,R.J., Simon,M., Slayman,C., Hunkapiller,M., Bolanos,R., Delcher,A., Dew,I., Fasulo,D., Flanigan,M., Florea,L., Halpern,A., Hannenhalli,S., Kravitz,S., Levy,S., Mobarry,C., Reinert,K., Remington,K., bu-Threideh,J., Beasley,E., Biddick,K., Bonazzi,V., Brandon,R., Cargill,M., Chandramouliswaran,I., Charlab,R., Chaturvedi,K., Deng,Z., Di,F., V, Dunn,P., Eilbeck,K., Evangelista,C., Gabrielian,A.E., Gan,W., Ge,W., Gong,F., Gu,Z., Guan,P., Heiman,T.J., Higgins,M.E., Ji,R.R., Ke,Z., Ketchum,K.A., Lai,Z., Lei,Y., Li,Z., Li,J., Liang,Y., Lin,X., Lu,F., Merkulov,G.V., Milshina,N., Moore,H.M., Naik,A.K., Narayan,V.A., Neelam,B., Nusskern,D., Rusch,D.B., Salzberg,S., Shao,W., Shue,B., Sun,J., Wang,Z., Wang,A., Wang,X., Wang,J., Wei,M., Wides,R., Xiao,C., Yan,C., Yao,A., Ye,J., Zhan,M., Zhang,W., Zhang,H., Zhao,Q., Zheng,L., Zhong,F., Zhong,W., Zhu,S., Zhao,S., Gilbert,D., Baumhueter,S., Spier,G., Carter,C., Cravchik,A., Woodage,T., Ali,F., An,H., Awe,A., Baldwin,D., Baden,H., Barnstead,M., Barrow,I., Beeson,K., Busam,D., Carver,A., Center,A., Cheng,M.L., Curry,L., Danaher,S., Davenport,L., Desilets,R., Dietz,S., Dodson,K., Doup,L., Ferriera,S., Garg,N., Gluecksmann,A., Hart,B., Haynes,J., Haynes,C., Heiner,C., Hladun,S., Hostin,D.,

Houck,J., Howland,T., Ibegwam,C., Johnson,J., Kalush,F., Kline,L., Koduru,S., Love,A., Mann,F., May,D., McCawley,S., McIntosh,T., McMullen,I., Moy,M., Moy,L., Murphy,B., Nelson,K., Pfannkoch,C., Pratts,E., Puri,V., Qureshi,H., Reardon,M., Rodriguez,R., Rogers,Y.H., Romblad,D., Ruhfel,B., Scott,R., Sitter,C., Smallwood,M., Stewart,E., Strong,R., Suh,E., Thomas,R., Tint,N.N., Tse,S., Vech,C., Wang,G., Wetter,J., Williams,S., Williams,M., Windsor,S., Winn-Deen,E., Wolfe,K., Zaveri,J., Zaveri,K., Abril,J.F., Guigo,R., Campbell,M.J., Sjolander,K.V., Karlak,B., Kejariwal,A., Mi,H., Lazareva,B., Hatton,T., Narechania,A., Diemer,K., Muruganujan,A., Guo,N., Sato,S., Bafna,V., Istrail,S., Lippert,R., Schwartz,R., Walenz,B., Yooseph,S., Allen,D., Basu,A., Baxendale,J., Blick,L., Caminha,M., Carnes-Stine,J., Caulk,P., Chiang,Y.H., Coyne,M., Dahlke,C., Mays,A., Dombroski,M., Donnelly,M., Ely,D., Esparham,S., Fosler,C., Gire,H., Glanowski,S., Glasser,K., Glodek,A., Gorokhov,M., Graham,K., Gropman,B., Harris,M., Heil,J., Henderson,S., Hoover,J., Jennings,D., Jordan,C., Jordan,J., Kasha,J., Kagan,L., Kraft,C., Levitsky,A., Lewis,M., Liu,X., Lopez,J., Ma,D., Majoros,W., McDaniel,J., Murphy,S., Newman,M., Nguyen,T., Nguyen,N., and Nodell,M. (2001). The sequence of the human genome. Science *291*, 1304-1351.

vidor-Reiss,T., Maer,A.M., Koundakjian,E., Polyanovsky,A., Keil,T., Subramaniam,S., and Zuker,C.S. (2004). Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. Cell *117*, 527-539.

Wade,R.H., Meurer-Grob,P., Metoz,F., and Arnal,I. (1998). Organisation and structure of microtubules and microtubule-motor protein complexes. Eur. Biophys. J. *27*, 446-454.

Walczak,C.E. and Nelson,D.L. (1994). Regulation of dynein-driven motility in cilia and flagella. Cell Motil. Cytoskeleton *27*, 101-107.

Wargo,M.J., Dymek,E.E., and Smith,E.F. (2005). Calmodulin and PF6 are components of a complex that localizes to the C1 microtubule of the flagellar central apparatus. J. Cell Sci. *118*, 4655-4665.

Warner,F.D. and Satir,P. (1974). The structural basis of ciliary bend formation. Radial spoke positional changes accompanying microtubule sliding. J. Cell Biol. *63*, 35-63.

Welsh,M.J. and Zabner,J. (1999). Cationic lipid mediated gene transfer of CFTR: safety of a single administration to the nasal epithelia. Hum. Gene Ther. *10*, 1559-1572.

Wemmer,K.A. and Marshall,W.F. (2004). Flagellar motility: all pull together. Curr. Biol. *14*, R992-R993.

Witman,G.B. (1992). Axonemal dyneins. Curr. Opin. Cell Biol. *4*, 74-79.

Witman,G.B. and Minervini,N. (1982). Dynein arm conformation and mechanochemical transduction in the eukaryotic flagellum. Symp. Soc. Exp. Biol. *35*, 203-223.

Wodehouse,T., Kharitonov,S.A., Mackay,I.S., Barnes,P.J., Wilson,R., and Cole,P.J. (2003). Nasal nitric oxide measurements for the screening of primary ciliary dyskinesia. Eur. Respir. J. *21*, 43-47.

Woolley,D. (2000). The molecular motors of cilia and eukaryotic flagella. Essays Biochem. *35*, 103-115.

Yang,P., Diener,D.R., Yang,C., Kohno,T., Pazour,G.J., Dienes,J.M., Agrin,N.S., King,S.M., Sale,W.S., Kamiya,R., Rosenbaum,J.L., and Witman,G.B. (2006). Radial spoke proteins of Chlamydomonas flagella. J. Cell Sci. *119*, 1165-1174.

Yokoyama,T., Copeland,N.G., Jenkins,N.A., Montgomery,C.A., Elder,F.F., and Overbeek,P.A. (1993). Reversal of left-right asymmetry: a situs inversus mutation. Science *260*, 679-682.

Yost,H.J. (2003). Left-right asymmetry: nodal cilia make and catch a wave. Curr. Biol. *13*, R808-R809.

Zariwala,M., Noone,P.G., Sannuti,A., Minnix,S., Zhou,Z., Leigh,M.W., Hazucha,M., Carson,J.L., and Knowles,M.R. (2001). Germline mutations in an intermediate chain dynein cause primary ciliary dyskinesia. Am. J. Respir. Cell Mol. Biol. *25*, 577-583.

Zariwala,M., O'Neal,W.K., Noone,P.G., Leigh,M.W., Knowles,M.R., and Ostrowski,L.E. (2004). Investigation of the possible role of a novel gene, DPCD, in primary ciliary dyskinesia. Am. J. Respir. Cell Mol. Biol. *30*, 428-434.

Zhang,M., Bolfing,M.F., Knowles,H.J., Karnes,H., and Hackett,B.P. (2004). Foxj1 regulates asymmetric gene expression during left-right axis patterning in mice. Biochem. Biophys. Res. Commun. *324*, 1413-1420.

Zhang,Q., Murcia,N.S., Chittenden,L.R., Richards,W.G., Michaud,E.J., Woychik,R.P., and Yoder,B.K. (2003). Loss of the Tg737 protein results in skeletal patterning defects. Dev. Dyn. *227*, 78-90.

Zhang,Y.J., O'Neal,W.K., Randell,S.H., Blackburn,K., Moyer,M.B., Boucher,R.C., and Ostrowski,L.E. (2002). Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia. J. Biol. Chem. *277*, 17906-17915.

Zito,I., Downes,S.M., Patel,R.J., Cheetham,M.E., Ebenezer,N.D., Jenkins,S.A., Bhattacharya,S.S., Webster,A.R., Holder,G.E., Bird,A.C., Bamiou,D.E., and Hardcastle,A.J. (2003). RPGR mutation associated with retinitis pigmentosa, impaired hearing, and sinorespiratory infections. J. Med. Genet. *40*, 609-615.

APPENDIX 1

Multipoint MERLIN linkage analysis for families 120, 130, 141, 143 and 145 following 600kb *Illumina* SNP screen.

Genetic linkage analysis for the region identified on chromosome 9

Haplotype analysis the region identified on chromosome 9

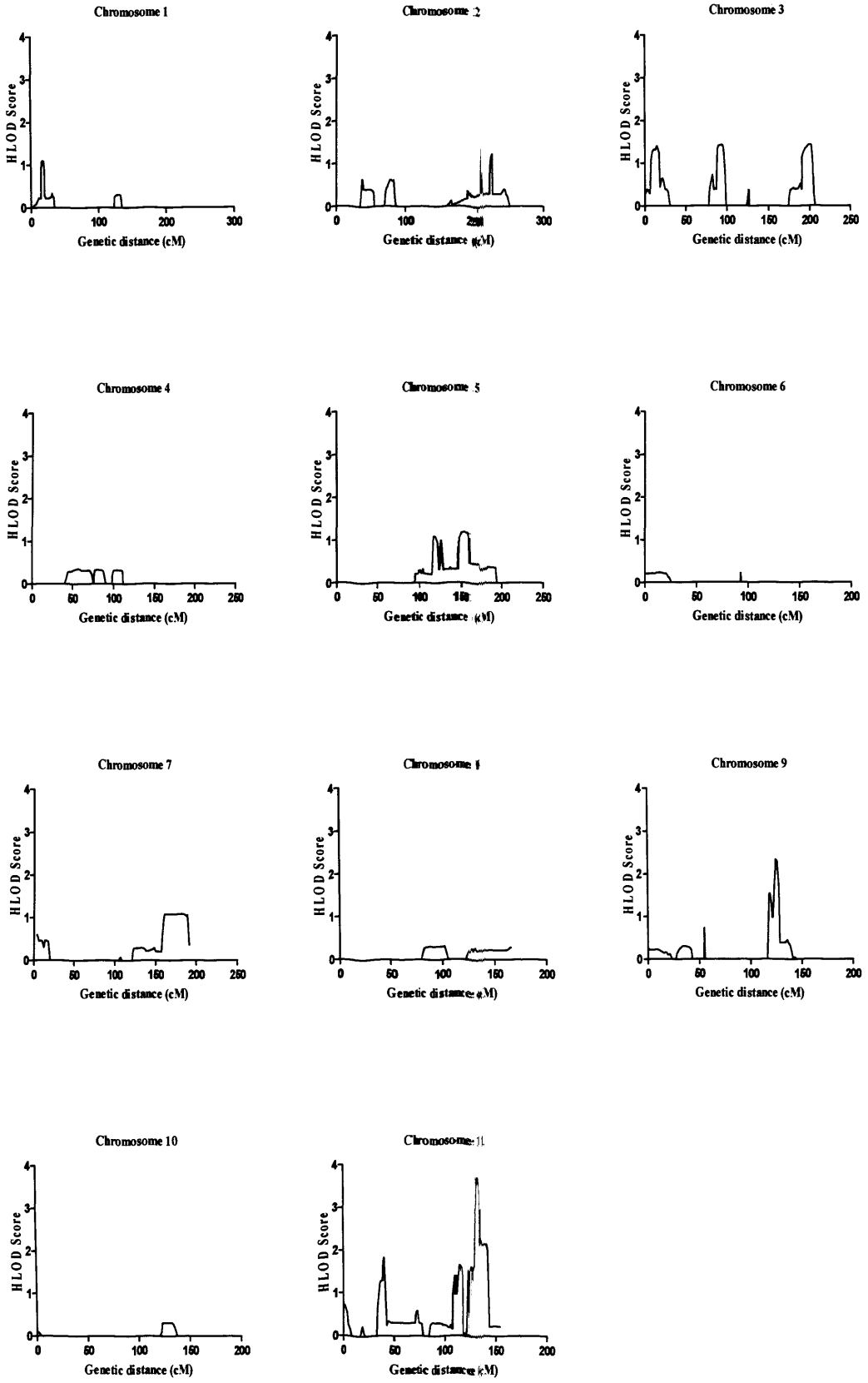
Novel in house markers used for fine mapping of the chromosome 17 locus

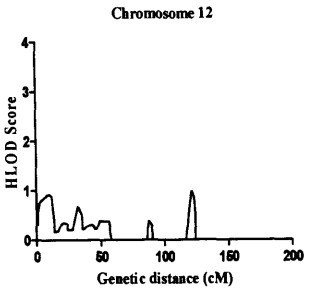
Comparative BLAST and function analysis of genes within the region of interest on chromosome 11 and 17

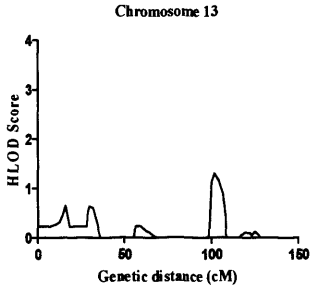
Primer sequences used for sequencing

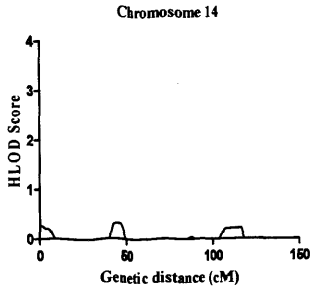
Candidate gene sequences with primers highlighted

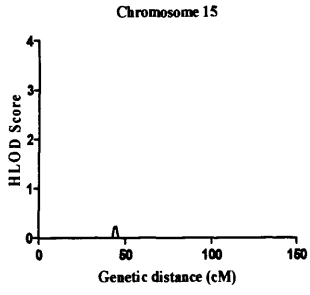
Multipoint MERLIN linkage analysis for families 120, 130, 141, 143 and 145 following 600kb *Illumina* SNP screen. HLOD scores >3 on chromosomes 11 and 17.

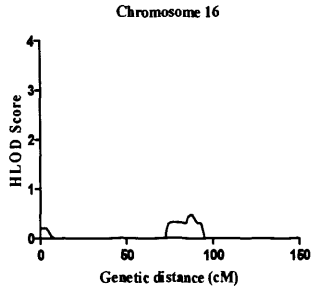


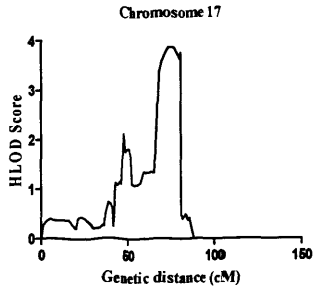

Chromosome 12
HLOD Score
Genetic distance (cM)

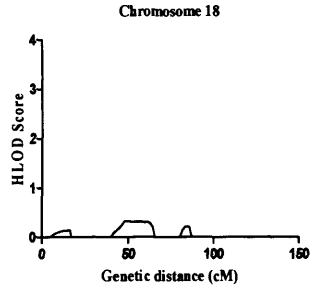

Chromosome 13
HLOD Score
Genetic distance (cM)

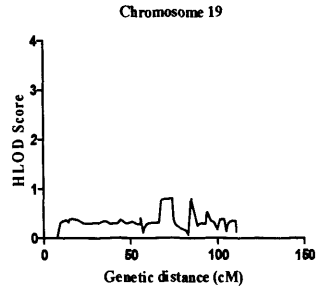

Chromosome 14
HLOD Score
Genetic distance (cM)

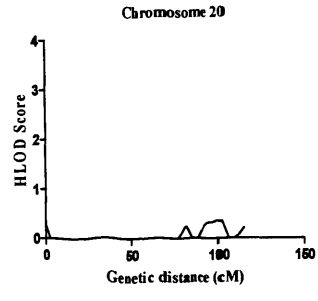

Chromosome 15
HLOD Score
Genetic distance (cM)

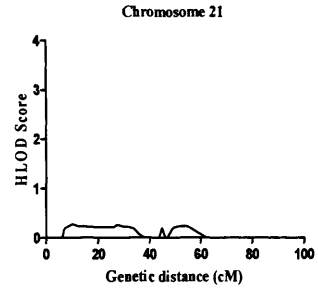

Chromosome 16
HLOD Score
Genetic distance (cM)

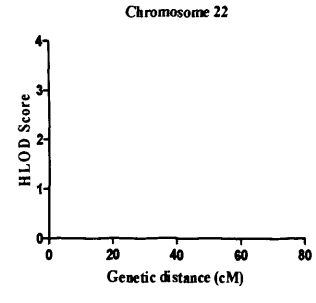

Chromosome 17
HLOD Score
Genetic distance (cM)


Chromosome 18
HLOD Score
Genetic distance (cM)


Chromosome 19
HLOD Score
Genetic distance (cM)

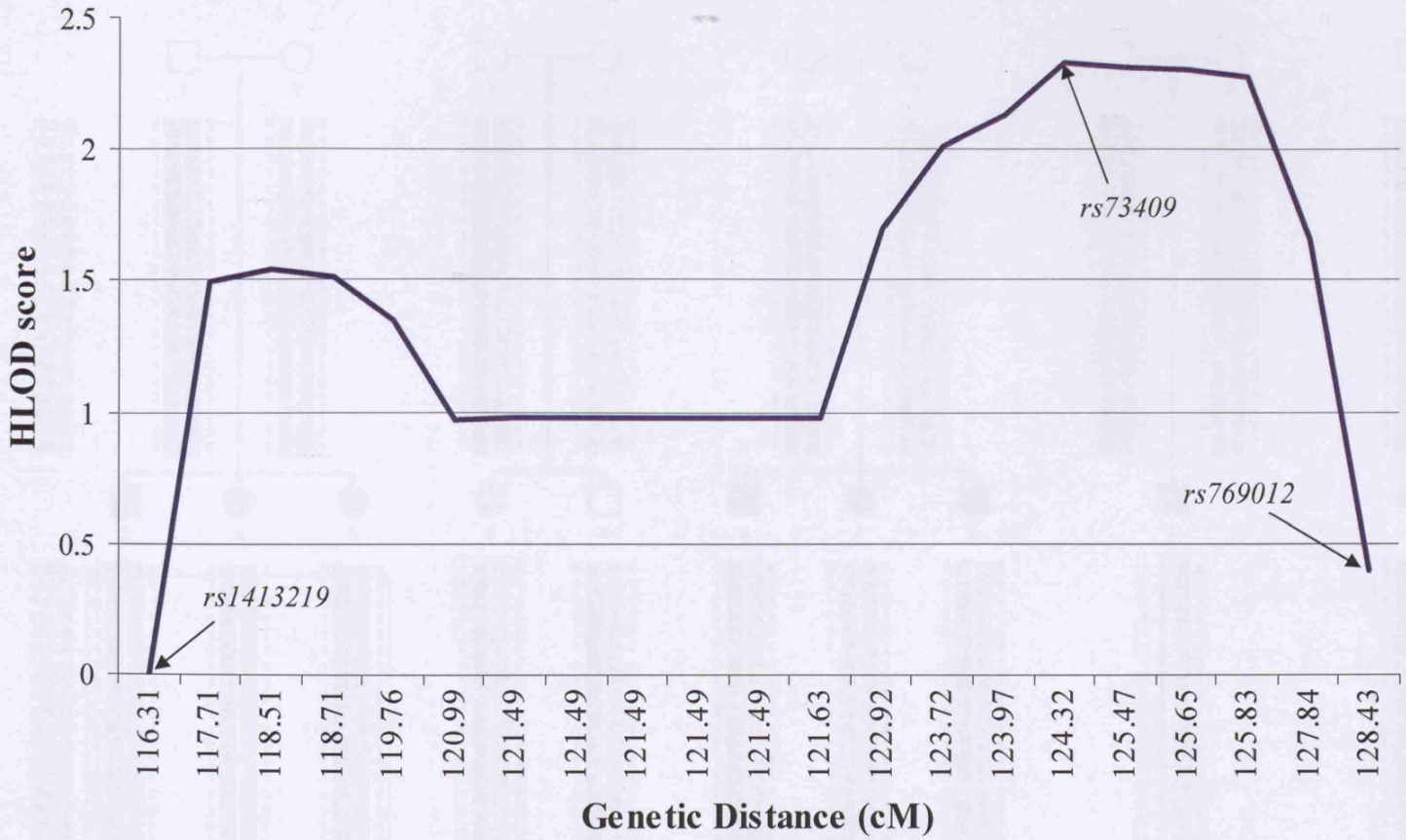

Chromosome 20
HLOD Score
Genetic distance (cM)


Chromosome 21
HLOD Score
Genetic distance (cM)

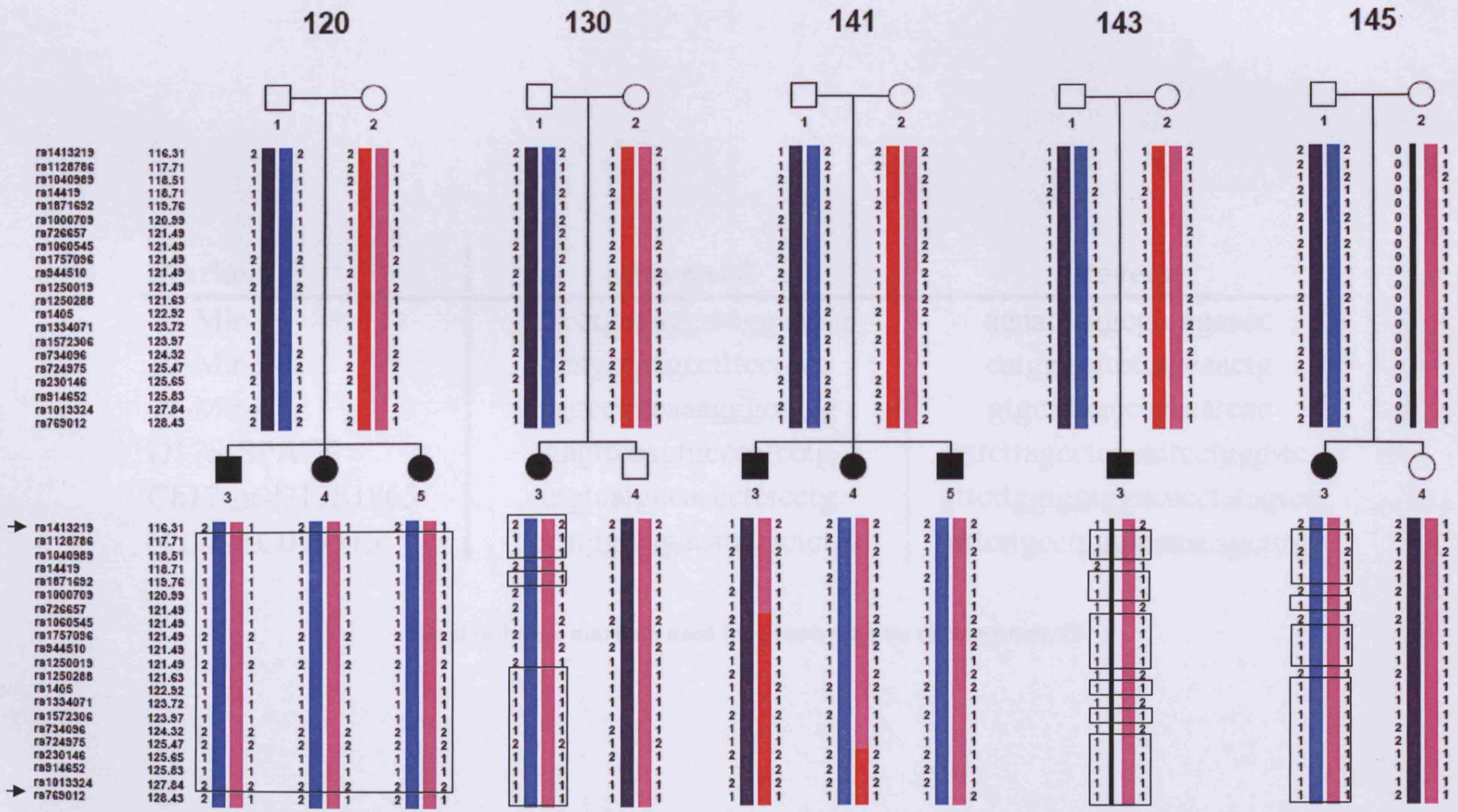

Chromosome 22
HLOD Score
Genetic distance (cM)

		Illumina screen SNPs only	All 5 families - Illumina screen SNPs plus extra microsatellite markers		
Marker name	**Genetic distance (cM)**	**LOD score**	**LOD**	**alpha**	**HLOD**
rs1413219	116.31	-0.041	-8.445	0	0
rs1128786	117.71	1.866	-3.805	0.418	1.492
rs1040989	118.51	1.976	-4.18	0.382	1.542
rs14419	118.71	1.988	-6.535	0.361	1.517
rs1871692	119.76	2.019	-6.062	0.335	1.351
rs1000709	120.99	2.039	-9.615	0.193	0.97
rs726657	121.49	2.044	-10.005	0.193	0.974
rs1060545	121.49	2.044	-10.005	0.193	0.974
rs1757096	121.49	2.044	-10.005	0.193	0.974
rs944510	121.49	2.044	-10.005	0.193	0.974
rs1250019	121.49	2.044	-10.005	0.193	0.974
rs1250288	121.63	2.044	-9.004	0.194	0.976
rs1405	122.92	2.046	-3.992	0.409	1.695
rs1334071	123.72	2.048	-3.265	0.474	2.013
rs1572306	123.97	2.048	-3.052	0.497	2.128
rs734096	124.32	2.047	-2.737	0.53	**2.334**
rs724975	125.47	2.028	-2.663	0.53	2.312
rs230146	125.65	2.022	-2.734	0.529	2.302
rs914652	125.83	2.009	-2.805	0.527	2.282
rs1013324	127.84	1.575	-3.134	0.467	1.653
rs769012	128.43	-0.038	-5.461	0.221	0.397

Genetic linkage analysis on chromosome 9. LOD, alpha and HLOD score shown for all five Mirpur PCD families after high resolution marker typing using *Illumina* SNP genotyping results. Genetic distance in centimorgans (cM) according to the *Illumina* genetic map constructed by deCODE. Maximum HLOD at this region highlighted in red.



HLOD plot across region of interest on chromosome 9 for all five Mirpur PCD families. Maximum HLOD of 2.3 achieved at marker *rs73409* (127.54 cM). Critical region defined by loss of homozygosity at marker *rs1413219* and *rs769012* amongst affected individuals in family 120.



High density haplotyping of Mirpur families at the putative locus on chromosome 9. Genetic distance in cM (centimorgans). Boxing indicates homozygosity. Arrows indicate flanking markers.

Marker	Forward	Reverse
17-Mir-1	ggcctgagaggataggatgag	agaatcaggcgtcggaaac
17-Mir-3	gatgctgagcctttcctcac	catgggtgtctgtgcaactg
17-Mir-4	tggcctatcaaagggctagg	gtgcacagccacacatcac
D17S-SPAG9	gagttcaagtgcccatcctg	gttcttagcctcgaattcctgggttc
Ch17-nr-D17S1865	atgtcatgccaccctctcctg	gttcttggtggtggacacctatagtcc
Ch17-AC019315	ccatggtaagacatagcctctc	gttcttgcctgggataaacagctttg

Novel in-house markers used for genotyping on chromosome 17

Comparative BLAST and function analysis of genes within the region of interest on chromosome 11, as described in Chapter 3, Section 3.1.7.1

Accession	Location (Mb)	Gene	ARAB	CERV	POMB	CHLY	PAZO	DEPLA	TRYP	CION	TETR	McKE	ELEG	fox1	Function	Domains identified by NCBI BLASTp	Unigene expression
NM_012101.2	119487205-119514073	tripartite motif-containing 29 (TRIM29),	0	0	0	0	0	0	1	1	0	0	0	1	It is able to complement the radiosensitivity defect of an ataxia telangiectasia (AT) fibroblast cell line	B-Box-type zinc finger; zinc binding domain	placenta, prostate and thymus
NM_178507.2	119586957-119605859	OAF homolog (Drosophila) (OAF),	0	0	0	0	0	0	0	0	0	0	0	0	Function unknown	No conserved domains detected	brain, nerve cord. expressed in embryonic, larval and adult gonads of both sexes, and larval imaginal disks
NM_014352.1	119616256-119695863	POU domain, class 2, transcription factor 3 (POU2F3),	0	1	0	0	0	0	0	1	0	0	1	0	POU2F3 in cervical cancer cells may be CC-related tumor suppressor genes	POU domain, Homeodomain	colon, esophagus, pancreas, placenta, testis, uterus
NM_174926.1	119701226-119706556	transmembrane protein 136 (TMEM136),	0	0	0	0	0	0	0	1	0	0	0	0	Function of protein unknown	TLC	placenta, prostate and thymus. Colocalizes with intermediate filaments
NM_015313.1	119713156-119865855	Rho guanine nucleotide exchange factor (GEF) 12 (ARHGEF12),	0	1	1	1	0	0	0	1	1	0	1	1	Rho GTPases play a fundamental role in numerous cellular processes that are initiated by extracellular stimuli that work through G protein coupled receptors. The encoded protein may form a complex with G proteins and stimulate Rho-dependent signals. This	PDZ or RhoGEF domain	brain, colon, kidney, liver, lung, ovary, pancreas, prostate, testis, thyroid,
NM_014619.2	120036238-120362179	glutamate receptor, ionotropic, kainate 4 (GRIK4),	1	0	0	1	0	0	0	1	0	0	1	0	L-glutamate acts as an excitatory neurotransmitter at many synapses in the central nervous system. The postsynaptic actions of Glu are mediated by a variety of receptors that are named according to their selective agonists	ANF_receptor, PBPb, Ligand-gated ion channel domain	testis, brain, uterus
NM_152715.2	120400040-120463333	leucine rich repeat containing 35 (LRRC35),	1	1	1	1	1	1	1	1	1	0	1	0	novel regulator of tubulin stability with overexpression causing depolymerization of microtubules and suppression resulting in an increase in the number of stable microtubules	Trypanosoma - dynein light chain, Anthocidaris crassispina - outer arm dynein light chain 2, Aedes aegypti - dynein light chain,Homo sapiens - axonemal dynein light chain	brain, kidney, liver, lung heart, uterus, testis
NM_005422.1	120478585-120596725	tectorin alpha (TECTA),	0	0	0	0	0	0	0	1	1	0	1	1	Alpha-tectorin is one of the major noncollagenous components of the tectorial membrane	NIDO, von Willebrand factor (vWF) type D domain, Trypsin Inhibitor like cysteine rich domain, Zona pellucida (ZP) domain	brain, testis, lung, uterus, kidney
NM_006918.3	120668708-120684556	sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like (SC5DL),	1	1	1	1	0	1	1	1	1	0	1	0	encodes an enzyme that is involved in cholesterol biosynthesis. It is thought to be an integral membrane protein	Sterol desaturase	brain, heart, lung, testis, thymus, uterus, liver, esophagus, kidney,
NM_003105.3	120828130-121005610	sortilin-related receptor, L(DLR class) A repeats-containing (SORL1),	0	1	1	1	0	0	0	1	1	0	1	1	multifunctional endocytic receptor, that may be implicated in the uptake of lipoproteins and of proteases.	lipoprotein receptor	brain, heart, kidney, liver, lung, ovary, testis, thymus, trachea
NM_001001786.1	121491272-121492133	BRCC2 (BRCC2),	0	0	0	0	0	0	0	0	0	0	0	0	proapoptotic molecule through the caspase-dependent mitochondrial pathway of cell death	none	brain, heart, kidney, liver, lung, testis, thymus, trachea
NM_032873.3	122031640-122190390	Cbl-interacting protein Sts-1 (STS-1),	0	0	1	1	0	0	0	1	0	0	1	0	inhibit endocytosis of epidermal growth factor receptor (EGFR) and platelet-derived growth factor receptor	peptidase and ubiquitin associated domains	brain, heart, kidney, liver, lung, testis, thymus,
NM_019604.2	122214485-122248557	class-I MHC-restricted T cell associated molecule (CRTAM),	0	0	0	0	0	0	0	1	0	0	1	0	regulates a large panel of cell/cell interactions both within and outside of the immune system	spermatogenic immunoglobulin superfamily	brain, kidney, lung, thymus
NM_024806.2	122258683-122335639	chromosome 11 open reading frame 63 (C11orf63),	0	0	0	0	0	0	0	1	1	0	0	0	unknown	none	brain, kidney, lung, testis, thymus

NM_006597.3	122433411-122438054	heat shock 70kDa protein 8 (HSPA8).	1	1	1	1	1	1	1	1	1	1	1	1	endogenous inhibitory regulator of HSC70 by competing the cochaperones	hsp70 domain (chaperone)	testis
NM_024769.2	122448245-122571217	adipocyte-specific adhesion molecule (ASAM).	0	0	0	0	0	0	0	1	0	0	1	1	adipocyte-specific adhesion molecule	coxsackie virus and adenovirus receptor	brain, heart, kidney, liver, lung, testis
NM_018400.3	123007133-123030522	sodium channel, voltage-gated, type III, beta (SCN3B).	0	0	0	0	0	0	0	1	0	0	0	1	Modulates channel gating kinetics.	Myelin protein, sodium channel voltage gated	brain, kidney, liver, lung, testis
NM_003455.2	123100207-123117573	zinc finger protein 202 (ZNF202).	1	1	1	0	0	0	1	1	1	0	1	1	Transciption factor, associated with cleft palate, but zinc finger so cant be ruled out due to relationship with ciliary genes.	krueppel associated box, zinc finger domain	brain, heart, kidney, lung, testis, thymus, trachea
NM_001013743.1	123258845-123261550	PMP22 claudin domain-containing protein (PMP22CD).	0	0	0	0	0	0	0	0	0	0	0	0	involved in the formation of tight junction strands.	none	testis
NM_001004474.1	123352613-123353608	olfactory receptor, family 10, subfamily S, member 1 (OR10S1).	0	0	0	0	0	0	0	1	0	0	1	0	Putative odorant receptor.	7TM domain	unknown
NM_014622.4	123491321-123522826	loss of heterozygosity, 11, chromosomal region 2, gene A (LOH11CR2A).	0	0	0	1	0	0	0	1	1	0	0	0	May play a role in tumorigenesis as a tumor suppressor.	trypsin inhibitor	brain, kidney, liver, lung, testis, thymus, trachea
NM_052959.2	123986663-123995459	pannexin 3 (PANX3).	0	0	0	0	0	0	0	0	0	0	0	0	Structural component of the gap junctions.	none	whole body
NM_032811.1	123997981-124009440	transforming growth factor beta regulator 1 (TBRG1).	1	0	0	1	0	0	0	1	1	0	0	0	GF regulator	Domain of unknown function, but occurs frequently in chromatin-associated proteins	brain, heart, kidney, liver, lung, testis, thymus, trachea
NM_170801.3	124010904-124048927	sialic acid acetylesterase (SIAE).	0	0	0	0	0	0	0	1	0	0	0	0	removes acetyl groups fromsialic acids. Note, gene opposite sp17	none	brain, heart, kidney, liver, lung, testis, thymus
NM_017425.2	124048950-124069895	sperm autoantigenic protein 17 (SPA17).	0	0	0	0	0	0	0	1	1	0	0	1	zona pellucida-binding protein, but its ubiquitous expression suggests a broader role in cell-cell adhesion and/or cell migration. Expressed in cilia and flagella - pose multifunctional role.	Chlamydomonas RSP11 &RSP7 - due to domains?	brain, kidney, liver, lung, testis, trachea
NM_006176.1	124114952-124122309	neurogranin (protein kinase C substrate, RC3) (NRGN).	0	0	0	0	0	0	0	0	0	0	0	0	postsynaptic protein kinase substrate - possible target for thyroid hormone in human brain	sp17 - C-terminus (calmodulin)	brain, kidney, lung, testis, thymus
NM_014312.3	124122581-124127319	V-set and immunoglobulin domain containing 2 (VSIG2).	0	0	0	1	0	0	0	1	0	0	1	1	V-set and immunoglobulin domain containing protein 2 precursor	Ig domain	heart, liver, lung, testis
NM_138961.1	124126242-124137385	endothelial cell adhesion molecule (ESAM).	0	0	0	0	0	0	0	1	0	0	1	1	Endothelial cell-selective adhesion molecule precursor.	Ig domain	brain, heart, kidney, liver, lung, testis, thymus
NM_024631.1	124141804-124175509	chromosome 11 open reading frame 61 (C11orf61).	0	0	0	0	0	0	0	0	0	0	0	0	unknown	none	brain, heart, kidney, liver, testis, thymus

NM_022370.2	124240492-124256572	roundabout, axon guidance receptor, homolog 3 (Drosophila) (ROBO3),	0	0	0	1	0	0	0	1	0	0	1	1	Involved during neural development in axonal navigation at the ventral midline of the neural tube. In spinal chord development plays a role in guiding commissural axons probably by preventing premature sensitivity to Slit proteins thus inhibiting Slit signaling through ROBO1 (By similarity). Required for hindbrain axon midline	Fibronectin type 3 domain and Immunoglobulin domain cell adhesion molecule	brain, heart, kidney, lung, testis, thymus uterus, prostate
NM_152722.3	124294299-124311156	hepatocyte cell adhesion molecule (FLJ25530),	0	0	0	0	0	0	0	1	0	0	1	1	the cytoplasmic domain of FLJ25530 is essential to its function on cell matrix interaction and cell motility. encodes an Ig-like transmembrane glycoprotein and is involved in cell adhesion and growth control. may be associated with carcinogenesis of hepatocytes	IGcam	brain, muscle, liver, eye
NM_001037558.1	124294356-124295783	associated with liver cancer (HEPN1),	0	0	0	0	0	0	0	0	0	0	0	0	suppresses cell growth and induces apoptosis in HepG2 cells	No conserved domains detected	brain, muscle, liver, eye
NM_025004.1	124329227-124380307	coiled-coil domain containing 15 (CCDC15),	1	1	1	0	0	0	1	1	1	0	1	0	dna-dependent rna polymerase catalyzes the transcription of dna into rna using the four ribonucleoside triphosphates as substrates	No conserved domains detected	brain, kidney, liver, testis, thymus
NM_198277.1	124438442-124464342	solute carrier family 37 (glycerol-3-phosphate transporter), member 2 (SLC37A2),	1	1	0	1	0	0	1	1	1	0	1	0	Auxiliary transport protein activity	UhpC Sugar phosphate permease, NarK Nitrate/nitrite transporter	brain, heart, kidney, liver, testis, trachea, lung
NM_022062.1	124706863-124808135	PBX/knotted 1 homeobox 2 (PKNOX2),	1	1	0	1	0	0	0	1	0	0	1	0	may interact with PBX proteins and play a role in tissue-specific regulation of transcription	myeloid ecotropic viral integration site	kidney, liver, lung, thymus,thyroid, brain, eye, heart, spleen, skin, testis
NM_005103.3	124820860-124871333	fasciculation and elongation protein zeta 1 (zygin I) (FEZ1),	0	0	0	0	0	0	0	1	0	0	1	0	May be involved in axonal outgrowth as component of the network of molecules that regulate cellular morphology and axon guidance machinery. Able to restore partial locomotion and axonal fasciculation to C.elegans unc-76 mutants in germ-line transformation experiments	no conserved domains detected	heart, kidney, liver, lung, brain, ovary, testis, thymus
NM_004879.3	124944508-124959784	etoposide induced 2.4 (EI24),	0	0	0	1	0	0	0	1	1	0	1	0	an immediate-early induction target of p53-mediated apoptosis	Etoposide-induced protein 2.4 domain	brain, heart, kidney, liver, lung, ovary, testis, thymus,thyroid, lymph, pancreas
NM_152713.2	124967987-124996162	STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae) (STT3A),	1	1	1	1	0	0	1	1	1	0	1	0	Involved in protein glycosylation. Either required for the assembly of the oligosaccharyl transferase (OTase) complex or required in substoichiometric amounts for OTase activity (By similarity)	Oligosaccharyl transferase STT3 subunit or COG1287	high levels in placenta, liver, muscle,and pancreas, and at very low levels in brain, lung, and kidney. expressed in skin
NM_001274.2	125001547-125030847	CHK1 checkpoint homolog (S. pombe) (CHEK1),	1	1	1	1	1	1	1	1	1	0	1	1	Required for checkpoint mediated cell cycle arrest in response to DNA damage or the presence of unreplicated DNA	Serine/threonine protein kinase domain, Serine/Threonine protein kinases catalytic domain	most abundant expression in thymus, testis, small intestine and colon.
NM_001612.4	125047440-125056152	acrosomal vesicle protein 1 (ACRV1),	0	0	0	1	1	0	1	1	0	0	0	1	arises within the acrosomal vesicle during spermatogenesis and may be involved in sperm-zona binding or penetration	Ly-6 antigen / uPA receptor -like domain	testis, brain
NM_138294.2	125121398-125124952	expressed in prostate and testis (PATE),	0	0	0	0	0	0	0	0	0	0	0	0	phospholipase A2 activity, calcium ion binding, lipid catabolism	No conserved domains detected	prostate, testis, brain
NM_212555.1	125151236-125153924	chromosome 11 open reading frame 38 (C11orf38),	0	0	0	0	0	0	0	0	0	0	0	0	Function Unknown	No conserved domains detected	prostate
NM_145014.1	125258719-125275749	hydrolethalus syndrome 1 (HYLS1),	0	0	0	0	0	0	0	1	0	0	1	0	Phenotype = Hydrolethalus syndrome	No conserved domains detected	testis, liver, brain, kidney, lung, thymus
NM_031307.2	125268591-125278315	pseudouridylate synthase 3 (PUS3),	1	1	1	1	0	0	1	1	1	0	1	0	Molecular Function Isomerase activity,Pseudouridylate synthase activity Biological Process TRNA processing	Pseudouridine synthase domain	brain, heart, lung, ovary, testis,

NM_013264.2	125279801-125296201	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25 (DDX25),	1	1	1	1	1	1	1	1	1	0	1	1	alteration of RNA secondary structure, such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly, It may serve to maintain testicular functions related to steroidogenesis and spermatogenesis	DEAD-box helicases, helicase superfamily c-terminal domain	brain, lung, testis
NM_016952.3	125335784-125438397	Cdon homolog (mouse) (CDON),	0	0	0	1	0	0	0	1	0	0	1	1	The long isoform may play a role in neural and glial cell adhesion in the developing embryo. the short isoform may be a more general cell adhesion molecule involved in other tissues and imaginal disk morphogenesis. vital for embryonic development. essential for septate junctions. septate junctions, which are the equivalent of vertebrates	Igcam, FN3 domains	brain, heart, kidney, lung, ovary, testis
NM_032795.1	125577519-125586751	RNA pseudouridylate synthase domain containing 4 (RPUSD4),	1	1	1	1	0	0	1	1	1	0	1	0	pseudouridylate synthase activity, RNA binding and RNA processing	Pseudouridine synthase	brain, heart, kidney, liver, lung, ovary, testis, thymus, pancreas
NM_024556.2	125586829-125637750	family with sequence similarity 118, member B (FAM118B),	0	0	0	0	0	0	0	0	0	0	0	0	Function unknown	SIR2-like	brain, heart, kidney, liver, lung, ovary, testis, thymus
NM_003139.2	125638043-125643960	signal recognition particle receptor ('docking protein') (SRPR),	1	1	1	1	0	0	1	1	1	0	1	0	Ensures, in conjunction with SRP, the correct targeting of the nascent secretory proteins to the endoplasmic reticulum membrane system	Signal recognition particle alpha subunit N-terminal, signal recognition particle (SRP), SRP54-type protein GTPase domain	brain, heart, kidney, liver, lung, ovary, testis, thymus, uterus
NM_017547.2	125644265-125653236	FAD-dependent oxidoreductase domain containing 1 (FOXRED1),	0	0	0	1	0	0	0	1	0	0	1	0	Molecular Function: Oxidoreductase activity Biological Process: Electron transport	Glycine/D-amino acid oxidases	liver, thymus, heart, brain, kidney, lung, ovary, testis
NM_148910.2	125656212-125668281	toll-interleukin 1 receptor (TIR) domain containing adaptor protein (TIRAP), transcript variant 2,	0	0	0	0	0	0	0	0	0	0	0	0	Adapter involved in the TLR4 signaling pathway in the innate immune response. Acts via IRAK2 and TRAF-6, leading to the activation of NF kappa-B, MAPK1, MAPK3 and JNK, resulting in cytokine secretion and the inflammatory response	Toll - interleukin 1 - resistance	brain, heart, kidney, liver, lung, testis, thymus
NM_014026.3	125678857-125720853	decapping enzyme, scavenger (DCPS),	0	1	1	1	0	0	0	1	1	0	1	0	m7GDP breakdown by DcpS should prevent misincorporation of methylated nucleotides in nucleic acids, The central histidine within the DcpS HIT motif is critical for decapping activity	Scavenger decapping enzyme (DcpS),	kidney, liver, brain, heart, lung, ovary, testis, thymus
NM_006278.1	125731306-125789741	ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (ST3GAL4),	1	0	0	0	0	0	0	1	0	0	0	1	It may catalyze the formation of sequences found in terminal carbohydrate groups of glycoproteins and glycolipids. It may be involved in the biosynthesis of the sialyl Lewis X determinant and the glycolysis pathway	Glycosyltransferase family 29 (sialyltransferase),	brain, heart, kidney, liver, lung, ovary, prostate, testis
NM_032531.2	125799685-126375615	kin of IRRE like 3 (Drosophila) (KIRREL3),	0	0	0	0	0	0	0	1	0	0	1	1	interact with the C terminus of podocin and is involved in ensuring size- and charge-selective ultrafiltration in the kidney	Immunoglobulin domain	kidney brain, heart, liver, lung, ovary, testis
NM_173579.1	126378015-126381163	proline rich 10 (PRR10),	0	0	0	0	0	0	0	0	0	0	0	0	Function Unknown	No conserved domains detected	testis
NM_005238.2	127837486-127897371	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) (ETS1),	0	0	0	0	0	0	0	1	0	0	1	1	Transcription factor	SAM_PNT, SAM / Pointed domain, erythroblast transformation specific domain	brain, heart, kidney, liver, lung, ovary, prostate, testis, thymus
NM_002017.2	128069193-128186325	Friend leukemia virus integration 1 (FLI1),	0	0	0	0	0	0	0	1	0	0	1	1	is a suppressor of collagen transcription in human skin in vivo. Overexpression of this protein affects cell growth and differentiation of K562 cells. Fli-1 and GATA-1 work together to activate the expression of genes associated with the terminal differentiation of megakaryocytes	SAM_PNT, SAM / Pointed domain, erythroblast transformation specific domain	brain, heart, kidney, liver, lung, ovary, prostate, testis, thymus
NM_000220.2	128213125-128242478	potassium inwardly-rectifying channel, subfamily J, member 1 (KCNJ1), transcript variant rom-k1,	0	0	0	1	0	0	0	1	0	0	1	0	an integral membrane protein and inward-rectifier type potassium channel. It is activated by internal ATP and probably plays an important role in potassium homeostasis. The encoded protein has a greater tendency to allow potassium to flow into a cell rather than out of a cell.	Inward rectifier potassium channel	kidney, liver, lung, trachea
NM_145013.1	128274678-128280802	chromosome 11 open reading frame 45 (C11orf45),	1	0	0	0	0	0	0	0	0	0	0	0	Function Unknown	No conserved domains detected	trachea, heart, testis, brain, placenta, kidney,

NM_022112.1	126310482-126318032	p53-regulated apoptosis-inducing protein 1 (P53AIP1),	0	0	0	0	0	0	0	0	0	0	0	0	regulates the mitochondrial apoptotic pathway	No conserved domains detected	brain, lung, ovary, thymus, uterus
NM_014715.2	128343052-128399222	Rho GTPase-activating protein (RICS),	1	1	1	0	0	0	1	1	1	0	1	1	may regulate dendritic spine morphology and strength by modulating Rho GTPase	RhoGAP	trachea, thymus, testis, liver, lung, kidney, brain
NM_003658.3	126751045-126827025	BarH-like homeobox 2 (BARX2),	1	0	0	0	0	0	0	1	0	0	1	0	Transcription Factor and is involved in controlling the expression patterns of cell adhesion molecules	homeodomain, Sec62,	kidney
NM_138788.2	129190951-129235108	transmembrane protein 45B (TMEM45B),	1	0	0	1	0	0	0	1	0	0	0	0	Function Unknown	DUF716	trachea, lung, kidney, liver
NM_006165.2	129239575-129268114	nuclear factor related to kappaB binding protein (NFRKB),	1	0	0	0	1	0	0	1	1	0	1	0	Molecular Function: Specific RNA polymerase II transcription factor activity TAS GOA/IPI Biological Process: Inflammatory response TAS GOA/IPI and Transcription from RNA polymerase II promoter	No conserved domains detected	liver, testis, lung, barin, kidney, thymus, trachea
NM_020228.2	129274812-129377940	PR domain containing 10 (PRDM10),	1	1	1	0	0	0	1	1	1	0	1	1	transcription factor that contains C2H2-type zinc-fingers, may be involved in the development of the central nerve system (CNS), as well as in the pathogenesis of neuronal storage disease.	No conserved domains detected	brain, liver, ovary, testis, thymus, trachea
NM_001007543.1	129377729-129380591	chromosome 11 open reading frame 37 (C11orf37),	0	0	0	0	0	0	0	1	0	0	0	0	Function Unknown	No conserved domains detected	brain
NM_001642.1	129445011-129519910	amyloid beta (A4) precursor-like protein 2 (APLP2),	0	0	0	1	0	0	0	1	0	0	1	0	May play a role in the regulation of hemostasis. The soluble form may have inhibitory properties towards coagulation factors	A4_EXTRA_KU	brain, heart, lung, liver, kidney, testis, thymus, trachea
NM_021978.3	129534933-129585468	suppression of tumorigenicity 14 (colon carcinoma) (ST14),	0	0	0	1	0	0	0	1	1	0	1	1	The protein encoded by this gene is an epithelial-derived, integral membrane serine protease, which has been shown to cleave and activate hepatocyte growth factor/scattering factor, and urokinase plasminogen activator, which suggest the function of this protease as an epithelial membrane activator for other proteases and latent	Trypsin-like serine protease, CUB domain, Low Density Lipoprotein Receptor Class A domain	pancreas, ovary, lung, heart, kidney, liver, thymus, trachea
NM_014155.3	129601782-129689717	BTB (POZ) domain containing 15 (BTBD15),	1	1	1	1	0	0	0	1	1	0	1	1	transcriptional regulator that acts as repressor or activator. binds, in-vitro, to nf-e2 binding sites. play important roles in coordinating transcription activation and repression by mafk	BTB/POZ domain, Zinc finger, COG5048	testis, thymus, kidney, brain, heart, liver, lung, trachea
NM_001039787.1	129689629-129778343	hypothetical protein LOC646363 (FLJ34521),	0	0	0	0	0	0	0	0	0	0	0	0	Function Unknown	No conserved domains detected	testis, kidney, lung, brain, ovary
NM_007037.3	129780425-129804098	ADAM metallopeptidase with thrombospondin type 1 motif, 8 (ADAMTS8),	0	0	0	0	0	0	0	1	0	0	1	1	Has anti-angiogenic properties	Reprolysin, ADAM Cysteine-Rich Domain, ADAM-TS Spacer 1 domain, Thrombospondin type 1 domain.	Highly expressed in adult and fetal lung, lower expression in brain, placenta, heart, stomach and fetal brain and kidney
NM_173560.1	130048061-130092457	chromosome 11 open reading frame 44 (C11orf44),	0	0	0	0	0	0	0	1	0	0	0	0	Function Unknown	No conserved domains detected	testis
NM_014758.1	130250985-130291572	sorting nexin 19 (SNX19),	0	0	1	0	0	0	0	1	0	0	1	0	May be involved in several stages of intracellular trafficking	PX domain	brain, heart, kidney, liver, lung, testis, thymus, trachea
NM_207432.1	131033417-131038080	chromosome 11 open reading frame 39 (C11orf39),	0	0	0	0	0	0	0	1	0	0	0	0	Function Unknown	No conserved domains detected	brain
NM_016522.2	131285922-131711925	neurotrimin (HNT),	0	0	0	1	0	0	0	1	0	0	1	1	Neural cell adhesion molecule and may promote neurite outgrowth and adhesion via a homophilic mechanism	Immunoglobulin domain cell adhesion molecule (cam)	brain, heart, kidney, liver, lung, testis, trachea
NM_002545.3	131790087-132318247	opioid binding protein/cell adhesion molecule-like (OPCML),	0	0	0	0	0	0	0	1	0	0	1	1	Binds opioids in the presence of acidic lipids; probably involved in cell contact and may have an accessory role in opioid receptor function	Immunoglobulin domain cell adhesion molecule (cam)	brain, heart, kidney, liver, lung, ovary, testis
NM_174927.1	133215727-133220602	spermatogenesis associated 19 (SPATA19),	0	0	0	0	0	0	0	0	0	0	0	0	May have a role in spermiogenesis	No conserved domains detected	testis, placenta
NM_001001873.3	133272818-133276487	hypothetical protein LOC283174 (LOC283174),	0	0	0	0	0	0	0	0	0	0	0	0	Function Unknown	No conserved domains detected	brain, liver, lung, testis, uterus

Comparative BLAST and function analysis of genes within the region of interest on chromosome 17, as described in Chapter 3, Section 3.1.7.1

Accession	Gene	ARAB	CERV	POMB	CHLY	PAZO	DEFLA	TRYP	CION	TETR	McKE	ELEG	FoxJ1	Function	Domains identified by NCBI BLASTp	UniGene expression
NM_173623.1	hypothetical protein FLJ35808 (FLJ35808),	0	0	0	1	0	0	1	1	1	0	1	1	Tubulin tyrosine ligase activity (catalyses ATP-dependent post-translational addition of a tyrosine to the carboxy terminal end of detyrosinated alpha-tubulin)	Tubulin-tyrosine ligase family	brain, colon, lung, stomach, testis
NM_152347.3	hypothetical protein FLJ40342 (FLJ40342),	0	0	0	1	1	0	0	1	0	1	0	1	calcium ion binding activity, in mcKeen, not in C. Elegans	Ca2+-binding protein (EF-Hand superfamily)	inc.brain, kidney, lung, testis, embryo
NM_003971.3	sperm associated antigen 9 (SPAG9), ,	0	0	0	0	0	0	0	1	1	0	1	1	scaffolding protein invd in activation of specific signaling pathways, abundantly expressed in testis (sperm acrosomal region)	Intermediate filament protein	inc.brain, kidney, lung, testis, embryo
NM_033413.2	hypothetical gene MGC16309 (MGC16309), (LRRC46)	0	1	1	1	1	1	1	1	1	1	1	1	in McKeen and ciliates	Leucine-rich repeat	brain, eye, lung, testis
NM_016835.1	microtubule-associated protein tau (MAPT), ,	0	0	0	1	0	0	1	1	0	1	1	0	enriched in axons/ invd in alzheimers/parkinsons disease.. -ve regn MT depolymerisation	tubulin-binding repeat	inc.brain, kidney, lung, testis, embryo
NM_001006607.1	c114 SLIT-like testicular protein (LOC474170),	0	0	0	0	0	0	1	1	0	1	1	0	unknown function	no putative conserved domains	inc.brain, kidney, lung, testis, embryo
NM_016428.2	ABI gene family, member 3 (ABI3),	0	0	1	0	0	0	1	1	0	1	1	0	inhibits ectopic metastasis of tumor cells as well as cell migration, but in mcKeen	SH3 domain	
NM_002476.2	myosin, light polypeptide 4, alkali; atrial, embryonic (MYL4), ,	1	1	1	1	1	0	1	1	1	1	1	0	myosin alkali light chain that is found in embryonic muscle and adult atria	EF-hand, calcium binding motif	
NM_006695.3	RaP2 interacting protein 8 (RPIP8),	0	0	0	0	0	0	0	1	0	0	0	0	small GTPase regulator activity/signal transduction	domain involved in Ras-like GTPase signaling	
NM_001466.2	frizzled homolog 2 (Drosophila) (FZD2),	0	0	0	0	0	0	0	1	0	0	1	0	receptors for Wnt signaling proteins, signal by activating calcium release from intracellular stores	Frizzled/Smoothened family membrane region - mediates hedgehog signalling. cysteine rich domain	
NM_006460.1	HMBA-inducible (HIS1),	0	0	0	0	0	0	0	1	0	0	0	0	expressed in smooth muscle cells, function unknown	no putative conserved domains	
NM_030802.2	C/EBP-induced protein (LOC81558),	0	0	0	0	0	0	0	1	0	0	0	0	C/EBP-induced protein	no putative conserved domains	
NM_002055.2	glial fibrillary acidic protein (GFAP),	1	1	1	1	1	1	1	1	1	1	1	1	one of the major intermediate filament proteins of mature astrocytes	Intermediate filament protein, Chromosome segregation ATPases	
NM_018509.2	hypothetical protein PRO1855 (PRO1855),	1	1	1	1	1	0	1	1	1	1	1	0	in McKeen	Leucine-rich repeat (LRR) protein [Function unknown]	

NM_152344.1	hypothetical protein FLJ30656 (FLJ30656),	0	1	0	0	0	0	0	1	0	0	1	0		no putative conserved domains	
NM_024032.2	hypothetical protein MGC3130 (MGC3130),	1	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_144608.1	hypothetical protein MGC39389 (FLJ32384),	0	0	0	0	0	0	0	1	0	0	0	0		no putative conserved domains	
NM_015443.2	hypothetical protein LOC284058 (LOC284058),	0	0	0	0	0	0	0	1	0	0	0	0		no putative conserved domains	
NM_017928.1	hypothetical protein FLJ20694 (FLJ20694),	0	0	0	0	0	0	0	1	0	0	0	0		no putative conserved domains	
NM_024107.1	hypothetical protein MGC3123 (MGC3123),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_144609.1	hypothetical protein FLJ31795 (FLJ31795),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_152343.1	hypothetical protein FLJ25414 (FLJ25414),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_152466.1	hypothetical protein FLJ25168 (FLJ25168),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_178542.3	hypothetical protein DKFZp762C2414 (DKFZp762C2414),	1	0	1	1	1	1	0	1	0	0	1	1		no putative conserved domains	
NM_024320.2	hypothetical protein MGC11242 (MGC11242),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_001004335.1	FLJ42842 protein (FLJ42842),	0	0	0	0	0	0	0	1	0	0	0	0		no putative conserved domains	
NM_014834.2	hypothetical LOC9884 (LOC9884),	0	1	0	1	0	0	0	1	0	1	1	0		no putative conserved domains	
NM_153229.1	hypothetical protein FLJ33318 (FLJ33318),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_052855.2	hypothetical protein MGC15396 (MGC15396),	0	0	0	0	0	1	0	1	0	0	1	0		ankyrin repeats - mediates protein-protein interactions	

NM_032595.1	protein phosphatase 1, regulatory subunit 9B, spinophilin (PPP1R9B),	1	1	1	1	0	0	1	1	1	1	1	1	regulatory subunit of protein phosphatase-1 catalytic subunit, highly enriched in dendritic spines, in McKeen	PDZ domain (invd in protein/protein interactions), Chromosome segregation ATPases	
NM_016632.1	ADP-ribosylation factor-like (LOC51326),	1	1	1	1	1	1	1	1	1	0	1	1		ARF-like small GTPases	
NM_000212.2	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61) (ITGB3),	0	0	0	0	0	0	1	1	1	0	1	0	mediates platelet aggregation by acting as a receptor for fibrinogen	Integrin beta subunits	
NM_002507.1	nerve growth factor receptor (TNFR superfamily, member 16) (NGFR),	0	0	0	1	0	0	1	1	1	0	0	0	Nerve growth factor receptor invd in neurogenesis	Tumor necrosis factor receptor (TNFR) domain	
NM_032133.2	MYCBP associated protein (MYCBPAP),	0	0	0	1	1	0	0	1	1	0	0	0		no putative conserved domains	
NM_025237.1	sclerosteosis (SOST),	0	0	0	0	0	0	0	0	0	0	0	0	family of bone morphogenetic protein (BMP) antagonists, assd with sclerosteosis	C-terminal cysteine knot-like (CTCK) domain	
NM_002722.2	pancreatic polypeptide (PPY),	0	0	0	0	0	0	0	0	0	0	0	0	involved in the regulation of exocrine pancreatic secretion and biliary tract motility	Pancreatic Hormone domain	
NM_004160.2	peptide YY (PYY),	0	0	0	0	0	0	0	0	0	0	0	0	inhibitor of gastric acid secretion, gastric emptying, digestive enzyme secretion by the pancreas, and gut motility	Pancreatic Hormone domain	
NM_032376.2	hypothetical protein MGC4251 (MGC4251),	0	0	0	0	0	0	0	0	0	0	0	0	positive regulation of I-kappaB kinase/NF-kappaB cascade	no putative conserved domains	
NM_138387.2	glucose 6 phosphatase, catalytic, 3 (G6PC3),	0	0	0	0	0	0	0	0	0	0	0	0	hydrolase activity	Acid phosphatase homologues	
NM_145663.1	Dbf4-related factor 1 (DRF1), ,	0	0	0	0	0	0	0	0	0	0	0	0	serine-threonine kinase which links cell cycle regulation to genome duplication	Zinc finger in DBF-like proteins, Protein kinase essential for the initiation of DNA replication	
NM_016438.2	CLST 11240 protein (CLST11240),	0	0	0	0	0	0	0	0	0	0	0	0		Hypoxia induced protein conserved region	
NM_001007532.1	saltohin (STH),	0	0	0	0	0	0	0	0	0	0	0	0	within intron 9 of the tau gene, assd with late-onset Alzheimer disease	no putative conserved domains	
NM_203400.1	similar to candidate mediator of the p53-dependent G2 arrest (LOC388394),	0	0	0	0	0	0	0	0	0	0	0	0		no putative conserved domains	
NM_014726.1	ProSAPiP2 protein (ProSAPiP2),	0	0	0	0	0	0	0	0	0	0	0	0	part of the interaction network in the TNF/NFKB pathway	no putative conserved domains	

NM_145255.2	mitochondrial ribosomal protein L10 (MRPL10), nuclear gene encoding mitochondrial protein, ,	0	0	0	0	0	0	0	0	0	0	0	0	Mitochondrial ribosome	Ribosomal protein L10	
NM_032391.2	small nuclear protein PRAC (PRAC),	0	0	0	0	0	0	0	0	0	0	0	0	may play a regulatory role in the nucleus.	no putative conserved domains	
NM_004123.1	gastric inhibitory polypeptide (GIP),	0	0	0	0	0	0	0	0	0	0	0	0	stimulates insulin secretion	Glucagon like hormones	
NM_153446.1	UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase (GALGT2),	0	0	0	0	0	0	0	0	0	0	0	0	lipid glycosylation	Glycosyl transferase	
NM_007225.1	neurexophilin 3 (NXPH3),	0	0	0	0	0	0	0	0	0	0	0	0	signaling molecules that resembles neuropeptides	no putative conserved domains	
NM_170685.1	tachykinin 4 (hemokinin) (TAC4),	0	0	0	0	0	0	0	0	0	0	0	0	regulates B-cell development	no putative conserved domains	
NM_174920.2	hypothetical protein LOC201191 (LOC201191),	0	0	0	0	0	0	0	0	0	0	0	0		Sterile alpha motif (SAM): Widespread domain in signalling and nuclear proteins (mediates homodimerisation)	
NM_000023.1	sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein) (SGCA),	0	0	0	0	0	0	0	0	0	0	0	0	part of dystrophin-glycoprotein complex, invd in stability of muscle fiber membranes and to the linking of the actin cytoskeleton to the extracellular matrix	Dystroglycan-type cadherin-like domains	
NM_194072.1	spermatid-specific linker histone H1-like protein (HILS1),	0	0	0	0	0	0	0	0	0	0	0	0	displays characteristics of a linker histone and may regulate gene transcription, DNA repair	no putative conserved domains	
NM_006807.3	chromobox homolog 1 (HP1 beta homolog Drosophila) (CBX1),	0	0	0	0	0	0	0	1	1	0	1	0	localized at heterochromatin sites, where it mediates gene silencing	Chromo Shadow Domain	
NM_002204.1	integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor) (ITGA3), transcript variant a,	0	0	0	1	0	0	0	1	0	0	1	0	forms integrin that interacts with extracellular-matrix proteins	beta-propellor repeats, FG-GAP repeat	
NM_175575.4	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2 (WFIKKN2),	0	0	0	1	0	0	0	1	0	0	1	0	may control the action of multiple types of proteases	WAP domain, follistatin domain, immunoglobulin domain, two tandem Kunitz domains, and an NTR domain	
NM_177441.1	hypothetical protein MGC3123 (MGC3123),	0	0	0	0	0	0	0	1	0	0	1	0	ubiquitination	no putative conserved domains	
NM_213607.1	similar to RIKEN 4933439F11 (LOC388389),	0	0	0	1	0	0	0	1	0	0	0	0	invd in pancreatic cancer?	no putative conserved domains	
NM_004382.2	corticotropin releasing hormone receptor 1 (CRHR1),	0	0	0	0	0	0	0	1	0	0	1	0	mediates stress responses.	HormR, Domain present in hormone receptors, 7 transmembrane receptor	

NM_030753.3	wingless-type MMTV integration site family, member 3 (WNT3),	0	0	0	0	0	0	0	1	0	0	1	0	may play a key role in some cancers through activation of the WNT-beta-catenin-TCF signaling pathway	WNT1, found in Wnt-1	
NM_001012511.1	golgi SNAP receptor complex member 2 (GOSR2), transcript variant C,	0	0	0	0	0	0	0	1	0	0	1	0	trafficking membrane protein, may be involved in familial essential hypertension	Vesicle transport v-SNARE protein	
NM_013351.1	T-box 21 (TBX21),	0	0	0	0	0	0	0	1	0	0	1	0	transcription factor that controls the expression of the hallmark Th1 cytokine	T-box DNA binding domain	
NM_138355.2	secernin 2 (SCRN2),	0	0	0	1	0	0	0	0	0	0	1	0	dipeptidase activity	Peptidase family U34	
NM_003204.1	nuclear factor (erythroid-derived 2)-like 1 (NFE2L1),	0	0	0	0	0	0	0	1	0	0	1	0	may play a role in the regulation of heme synthesis and ferritin genes	basic region leucin zipper, bZIP Maf transcription factor	
NM_002144.2	homeo box B1 (HOXB1),	0	0	0	0	0	0	0	1	0	0	1	1	conserved family of transcription factors, exact role of this one not determined	Homeodomain	
NM_001934.2	distal-less homeobox 4 (DLX4), ,	0	0	0	0	0	0	0	1	0	0	1	0	postulated to play a role in forebrain and craniofacial development	Homeodomain	
NM_000088.2	collagen, type I, alpha 1 (COL1A1),	0	0	0	0	0	0	0	1	0	0	1	0	encodes the major component of type I collagen	Fibrillar collagens C-terminal domain	
NM_017643.1	mbt domain containing 1 (MBTD1),	0	0	0	0	0	0	0	1	0	0	1	0	chromatin modification, regn of transcription	MBT, involved in transcriptional regulation	
NM_145273.2	triggering receptor expressed on myeloid cells 4 (TREM4),	0	0	0	0	0	0	0	1	0	0	0	0	receptor activity	Immunoglobulin like	
NM_025104.2	Dbf4-related factor 1 (DRF1), ,	0	0	0	0	0	0	0	1	0	0	0	0	serine-threonine kinase which links cell cycle regulation to genome duplication	Zinc finger in DBF-like proteins, Protein kinase	
NM_005497.1	gap junction protein, alpha 7, 45kDa (connexin 45) (GJA7),	0	0	0	0	0	0	0	1	0	0	0	0	form gap junction channels	Connexin (participate in the regulation of signaling between developing and differentiated cell types)	
NM_006688.3	complement component 1, q subcomponent-like 1 (C1QL1),	0	0	0	0	0	0	0	1	0	0	0	0	phosphate transport	Complement component C1q domain, found in many collagens	
NM_148887.1	mitochondrial ribosomal protein L10 (MRPL10), nuclear gene encoding mitochondrial protein, ,	0	0	0	0	0	0	0	1	0	0	0	0	Mitochondrial ribosome	Ribosomal protein L10	
NM_031498.1	guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 2 (GNGT2),	0	0	0	0	0	0	0	1	0	0	0	0	play a crucial role in cone phototransduction	G protein gamma subunit-like motifs	

NM_001007529.1	FLJ40194 protein (FLJ40194),	0	0	0	0	0	0	0	1	0	0	0	0		no putative conserved domains	
NM_148571.1	mitochondrial ribosomal protein L27 (MRPL27), nuclear gene encoding mitochondrial protein, ,	0	0	0	0	0	0	0	1	0	0	0	0	Mitochondrial ribosomes	Ribosomal L27 protein	
NM_152463.1	essential meiotic endonuclease 1 homolog 1 (S. pombe) (EME1),	0	0	0	0	0	0	0	1	0	0	0	0	ATPase/H+ transport	no putative conserved domains	
NM_005749.2	transducer of ERBB2, 1 (TOB1),	0	0	0	0	0	0	0	0	0	0	1	0	inhibits T cell proliferation and transcription of cytokines and cyclins	anti-proliferation domains	
NM_001012479.1	granulin (GRN), ,	1	0	0	1	0	0	1	1	1	0	1	0	regulate cell growth - important in normal development, wound healing, and tumorigenesis	Granulin	
NM_175882.1	intramembrane protease 5 (IMP5),	1	0	0	1	0	0	1	1	1	0	1	1	proteolysis	Presenilin signal peptide peptidase, Protease associated domain	
NM_178500.2	phosphatase, orphan 1 (PHOSPHO1),	1	0	0	1	0	0	1	1	1	0	1	1	phosphatase	Putative Phosphatase, Phosphoserine phosphatase, Predicted hydrolase	
NM_198397.1	calcium channel, voltage-dependent, alpha 1G subunit (CACNA1G), 5,	1	0	0	1	0	0	1	1	1	0	1	0	Voltage-activated calcium channel	Ion transport protein	
NM_020178.3	carbonic anhydrase X (CA10),	1	0	0	1	0	0	1	1	1	0	1	0	zinc metalloenzymes, invd in reversible hydration of carbon dioxide	carbonic anhydrase domain	
NM_014798.1	pleckstrin homology domain containing, family M (with RUN domain) member 1 (PLEKHM1),	1	0	0	0	0	0	1	1	1	0	1	0	invd in intracellular signalling	Protein kinase C conserved region 1, PH domain, RUN domain	
NM_054022.2	golgi SNAP receptor complex member 2 (GOSR2), transcript variant B,	1	0	0	1	0	0	0	1	1	0	1	0	trafficking membrane protein, may be involved in familial essential hypertension	V-SNARE	
NM_176096.1	CDK5 regulatory subunit associated protein 3 (CDK5RAP3), transcript variant 3,	1	0	0	0	0	0	1	1	1	0	1	0	involved in neuronal differentiation	Protein of unknown function, poss CDK5 activator-binding domain	
NM_003563.3	speckle-type POZ protein (SPOP), ,	1	0	0	1	0	0	0	1	1	0	1	1	mediate transcriptional repression and interacts with components of histone deacetylase co-repressor complexes	BTB/POZ domain - mediates dimerisation	
NM_001002909.1	KIAA0553 protein (KIAA0553),	1	0	0	1	0	0	1	0	0	0	0	0	ATPase invd in DNA repair	ATPase	
NM_002390.2	a disintegrin and metalloproteinase domain 11 (ADAM11), ,	0	0	1	0	0	0	0	1	0	0	1	0	invd in cell-interactions such as fertilization, muscle development, and neurogenesis	ADAM Cysteine-Rich Domain, Reprolysin (M12B) family zinc metalloprotease	

NM_022167.1	xylosyltransferase II (XYLT2),	1	0	0	0	0	0	0	1	0	0	1	0	transfers xylose from UDP-xylose to specific serine residues of the core protein	Core-2/I-Branching enzyme	
NM_022827.2	sperm protein SSP411 (SSP411),	1	0	0	0	0	0	1	0	0	0	1	0	hydrolase activity on glycosyl bonds	thioredoxin domain	
NM_018346.1	radical S-adenosyl methionine domain containing 1 (RSAD1),	1	0	0	0	0	0	0	0	1	0	0	0	porphyrin biosynthesis	oxidoreductase domains	
NM_001267.1	chondroadherin (CHAD),	1	1	0	1	1	0	0	1	1	0	1	1	cartilage matrix protein	Leucine rich repeat	
NM_014233.1	upstream binding transcription factor, RNA polymerase I (UBTF),	1	1	0	0	0	0	1	1	1	0	1	1	transcription factor required for expression of the 18S, 5.8S, and 28S ribosomal RNAs	High Mobility Group (HMG)-box (DNA binding)	
NM_003726.2	src family associated phosphoprotein 1 (SCAP1),	0	1	1	0	0	0	1	1	0	0	1	1	src family kinase invd in T-cell activation	Src homology 3 domain, PH domain	
NM_153006.1	N-acetylglutamate synthase (NAGS),	0	1	1	0	0	0	0	0	0	0	0	0	mitochondrial enzyme that catalyzes the formation of N-acetylglutamate, not in ciliates		
NM_080863.4	ankyrin repeat and SOCS box-containing 16 (ASB16),	1	1	1	1	1	1	1	1	1	1	1	1	couple suppressor of cytokine signalling (SOCS) proteins and their binding partners with the elongin B and C complex, possibly targeting them for degradation	ankyrin repeats	
NM_000269.2	non-metastatic cells 1, protein (NM23A) expressed in (NME1), ,	1	1	1	1	1	1	1	1	1	1	1	1	regn of cell cycle	NDK - catalyze nonsubstrate specific conversions of nucleoside diphosphates to nucleoside triphosphates	
NM_003954.1	mitogen-activated protein kinase kinase kinase 14 (MAP3K14),	1	1	1	1	1	1	1	1	1	0	1	1	serine/threonine protein-kinase, stimulates NF-kappaB activity	Serine/Threonine protein kinases	
NM_005374.3	membrane protein, palmitoylated 2 (MAGUK p55 subfamily member 2) (MPP2),	1	1	1	1	1	0	1	1	1	0	1	0	interact with the cytoskeleton and regulate cell proliferation, signaling pathways, and intracellular junctions	SH3 domain, PDZ domain	
NM_004247.2	U5 snRNP-specific protein, 116 kD (U5-116KD),	1	1	1	1	1	0	1	1	1	0	1	1	splicing factor	Elongation factor Tu GTP binding domain	
NM_006178.1	N-ethylmaleimide-sensitive factor (NSF),	1	1	1	1	1	0	1	1	1	0	1	1	ATP-binding, proteolysis	AAA-superfamily of ATPases	
NM_002634.2	prohibitin (PHB),	1	1	1	1	0	0	1	1	1	1	1	1	negative regulator of cell proliferation and may be a tumor suppressor, mutations linked to sporadic breats cancer	prohibitin homologues	
NM_020038.1	ATP-binding cassette, sub-family C (CFTR/MRP), member 3 (ABCC3), transcript variant MRP3B,	1	1	1	1	1	0	1	1	1	0	1	1	member of the superfamily of ATP-binding cassette (ABC) transporters. may play a role in the transport of biliary and intestinal excretion of organic anions.	ATP-binding cassette	

NM_004090.2	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related) (DUSP3),	1	1	1	1	0	0	1	1	1	0	1	1	dephosphorylation	Dual specificity phosphatases	
NM_005474.3	histone deacetylase 5 (HDAC5), ,	1	1	1	1	0	0	1	1	1	0	1	1	chromatin modification, regn of transcription. Linked to colon cancer	Histone deacetylase family	
NM_016016.1	CGI-69 protein (CGI-69),	1	1	1	1	0	0	1	1	1	0	1	1	transport?	Mitochondrial carrier protein	
NM_024819.3	hypothetical protein FLJ22955 (FLJ22955),	1	1	1	1	0	0	1	1	1	0	1	1			
NM_021079.3	N-myristoyltransferase 1 (NMT1),	1	1	1	1	0	0	1	1	1	0	1	1			
NM_133373.2	phospholipase C, delta 3 (PLCD3),	1	1	1	1	0	0	1	1	1	0	1	1			
NM_024722.1	acyl-Coenzyme A binding domain containing 4 (ACBD4),	1	1	1	1	0	0	1	1	1	0	1	1			
NM_005892.3	formin-like 1 (FMNL1),	1	1	1	1	0	0	1	1	1	0	1	1	invd in limb formation-Shh pathway		
NM_001256.2	cell division cycle 27 (CDC27),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_006310.2	aminopeptidase puromycin sensitive (NPEPPS),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_002265.4	karyopherin (importin) beta 1 (KPNB1),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_145798.2	oxysterol binding protein-like 7 (OSBPL7), ,	1	1	1	1	0	0	1	1	1	0	1	0			
NM_016429.1	coatomer protein complex, subunit zeta 2 (COPZ2),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_005831.3	nuclear domain 10 protein (NDP52),	1	1	1	0	0	0	1	1	1	1	1	0			
NM_023079.2	hypothetical protein FLJ13855 (FLJ13855),	1	1	1	1	0	0	1	1	1	0	1	0	Ubiquitin-conjugating enzyme E2		

NM_007241.2	EAP30 subunit of ELL complex (EAP30),	1	1	1	1	0	0	1	1	1	0	1	0	SU of ELL complex (ribosomal TF)		
NM_005827.1	solute carrier family 35, member B1 (SLC35B1),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_007067.3	MYST histone acetyltransferase 2 (MYST2),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_025149.3	hypothetical protein FLJ20920 (FLJ20920),	1	1	1	1	0	0	1	1	1	0	1	0		Acyl-CoA synthetases, AMP-binding enzyme	
NM_017957.1	epsin 3 (EPN3),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_198396.1	calcium channel, voltage-dependent, alpha 1G subunit (CACNA1G), transcript variant 3,	1	1	1	1	0	0	1	1	1	0	1	0			
NM_016001.1	WD repeat domain 50 (WDR50),	1	1	1	1	0	0	1	1	1	0	1	0			
NM_001002027.1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1 (ATP5G1), nuclear gene encoding mitochondrial protein, ,	1	1	1	1	0	0	1	1	0	0	1	0			
NM_006546.3	IGF-II -binding protein 1 (IMP-1),	1	1	1	1	0	0	0	1	1	0	1	0			
NM_002611.3	pyruvate dehydrogenase kinase, isoenzyme 2 (PDK2),	1	1	1	1	0	0	1	1	0	0	1	0			
NM_006107.2	cisplatin resistance-associated overexpressed protein (CROP), ,	1	1	1	1	0	0	0	1	1	0	1	0			
NM_016424.3	cisplatin resistance-associated overexpressed protein (CROP), ,	1	1	1	1	0	0	0	1	1	0	1	0			
NM_000342.1	solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3, Diego blood group) (SLC4A1),	1	1	1	1	0	0	0	1	0	0	1	0			
NM_199262.2	Sp6 tranacription factor (SP6),	1	1	1	0	0	0	0	1	1	0	1	0			
NM_003110.4	Sp2 transcription factor (SP2),	1	1	1	0	0	0	0	1	1	0	1	0			

NM_018129.1	pyridoxine 5'-phosphate oxidase (PNPO),	1	1	1	1	0	0	0	1	0	0	1	0			
NM_013323.2	sorting nexin 11 (SNX11), ,	1	1	1	1	0	0	0	1	0	0	1	0	involved in intracellular trafficking	PhoX homologous domain of unknown function, PX domains bind to phosphoinositides	
NM_014897.1	zinc finger protein 652 (ZNF652),	1	1	1	0	0	0	0	1	1	0	1	0			
NM_199282.1	Rho GTPase activating protein 27 (ARHGAP27),	1	1	1	0	0	0	0	1	0	0	1	0			
NM_016504.2	mitochondrial ribosomal protein L27 (MRPL27), nuclear gene encoding mitochondrial protein, ,	1	1	1	0	0	0	0	1	1	0	0	0			

Exon	Fwd	Rev	size (bp)
1	CTTGGGTTGTGTTAGGGAAA	CTGCAGAATGGAGATGGTCTT	403
2	TTGGACTGCCAGTAAAGAACC	TGTGATGCCTCCTCACATAA	297
3	GTGGTACACACATTCCAAAAGC	AAGAGAGGCATGGCTGAAAA	532
4	CAGCTCTCAGATAGTTGCATCC	AGAAGGTTGATGGATTTGGA	314

Primer sequences for *SPA17* sequencing

Exon	Fwd	Rev	size (bp)
1	GTCTTGCAATGCCTAGTCTAG	GCAAACAAGGCAACACTGTG	337
2	TTCAGTACCCTTTGGTCCCA	GACACAATGGTTATTGCTGC	351
3	ACAGCTCTGATTACTGCCTG	GGAGATCAGCAGTACTCCTA	304
4	CCACAATTCCTAGCAGAGTC	GCCAATCATAGGCATCTGCA	525
5	TGCTCTTGCCTTGGATGAAG	TGCACTCCCACTAAACAGTG	367
6	ACTGGGCCACAAGCCAATTA	TTAACTGGCATCACCACCCT	255
7	GTCCACAGTACATGTGCTAG	CCCTGCAACAAGTCCTTATG	424

Primer sequences for *LRRC35* sequencing

Exon	Fwd	Rev	size (bp)
1	GGCACAGAATGGAAAGGTTC	TGAACAGAGGAAGAGGACAGG	179
2	GGAGTTTCTGGTTTCCAGCAT	CAATTTAGGTGAGCCAGGAG	400
3	CTGAGGTGGGCTTTGGTAGT	CTTATTGGGAGCAGGGGAGT	338
4	CATATCCCTTCCAGGCTTTG	GCCAGTCAGCCAATAGATCA	325
5	CATGTCCATCCCTCAAGGAG	ACAGCCTTTCGGTGTGTCTG	367
6	GCAGATGAATCTGGAAGTGGA	AAAGGGTCACAGCAAGTTGG	434
7	GCATAGGAAGGAACCCTTTG	CCCAAAGATTAGGACTGGTTC	506
8	CACCAGGAATTTGTGCTGAA	AAACACAGAGCACCTGCCAAC	387

Primer sequences for *FLJ35808* sequencing

Exon	Fwd	Rev	size (bp)
1 & 2	TCCTGAGTTCTGCCATCCTT	GGTCATCCTGAGTGTTTCAGC	358
4	TCCCTGCTAAGAGACCACAT	CTGGTTGGCTCCTTTCAGAT	247
6	AGACATGGTCTTGGGGCTTT	CCAGGCTTGGAGAAATCTGA	225
7	AGGTGCCCTCCTCTTTTCTT	CAGGGGCATGGCTAATAGAA	281
8	ATTCCCATCGCTTCATGGTC	ATGCAAAGACACGCAC	319
9	GTGCGTGTCTTTGCATCCT	GCCACCTTAGCTTCCCTCAT	289
10	TCCTTCCCACAAGGGACAA	TTTACTGGGGATGGGCTTCT	576

Primer sequences for *LRRC46* sequencing

SPA17 exon sequences with primers for sequencing highlighted.

Key

PRIMER
CODING EXON
START/STOP CODON
5'/3' UTR
INTRON
SNP

Exon 1
CATATTGAAAATTATGGGTTTTTTGCAGGATTTTAAATGATTTGGCCTAGTGAATGGGTCT
GTGTTTGTATTATACTAATTCACAATGGAGATTTCAAAACTCTTTGGAAGAAATGGATGGG
CTTGGGTTGTGTTAGGGAAAAAAGAATTCAAGTCCAGATGTATGAGTCTCAGAAATAATT
TGTTGGCATAGTTGTGATTATATAAGTAAAATCAGACAAATTTAAAGACAGTACTCTTTAA
TGAATCTTTAGGTTCCATAGGCAGTTCTTACCAAGAAGATGTCGATTCCATTCTCCA
ACACCCACTACCGAATTCCACAAGGATTTGGGAATCTTCTTGAAGGGCTGACACGCG
AGATTCTGAGAGAGCAACCGGACAATATACCAGCTTTTGCAGCAGCCTATTTTGAGA
GCCTTCTAGAGAAAAGAGAGAGTAAGCTTTCTAAAATTAGTCATTTTTTAAAATAAGAA
ACTAAGCATTTGTTTATGGTAAAACTTGCTTAAGACCATCTCCATTCTGCAGAAAGCCTTT
TATGAGGCAGGGAGTTTCAGATAACAGTTCTTGCTCTTTGAAACAGTATGTACCTCTTTGA
TATGT

Exon 2
ACTGGTAATGTCTGCAGTGCTACGCTATGCACAATGTCAATAAGATTTTGGACTGCCAGTA
AAGAACCTGTAATTTGAAACAAGAAACTTACATTTGTGTCACTATTCTACATTTACCTTAA
TGAATTCAAATAAATCTGAAGTTCTTATATTTTTATTATTTAGAAACCAACTTTGATCCA
GCAGAATGGGGGAGTAAGGTAGAAGACCGCTTCTATAACAATCATGCATTCGAGGTA
TGGTCCTTTGAAGCTGTTGGATTTGGCTATTTCTTCTCTCTGTCTACAGCTAGCAATTATT
AGGCTATCCCCAGAATTGTGAGAATTATGTGAGGAGGCATCACAAATAACTTATTAAATC

Exon 3
AAAACAGTGGTACACACATTCCAAAAGCAACAGATCAAACATTTTTGCCAATTATTTGGT
CTTTAATATTTGTTTACACGTATTTCCTTAAGAGCTTTTATTATTTCTTCTTGTTTCAATGTA
TTTTCCTTATTAAAAGATAAATACGTAGTATTTAATTCCCCCTATTTCTCAGTAATTTAAAG
TAATGTTTCATTACTTCTAAGCAGTTTGATACAAGACTTTGCCATTTTGGAAACATTGCTA
ATTGTGGCTCTCTCCTATCTGTCCAGGAGCAAGAACCACCTGAGAAAAGTGATCCTAAA
CAAGAAGAGTCTCAGATATCTGGGAAGGAGGAAGAGACATCAGTCACCATCTTAGTA
TGTAATATTTTTCATGTACTATTGGATTCCTACAAAGAATGATATTGCCCAATTTAAAACT
GAACTCCAAATATGAATGTCTGTTATCTTTATTGTCAAAATTATGACTTTACTTTCCAGAT
AATTAACATCAGTGATAAACTAACTGGTAGATAATTTTCAGCCATGCCTCTCTTGA

Exon 4
AAGCAACAGCTCTCAGATAGTTGCATCCCTAAAATGATTAAATTTCATCCATTTATTTTTG
TGGCATCAAAACTACGCCATATTTAAAGAGTCTCTTACTTTTTCTCCTTTCGATGAAGGAC
TCTTCTGAGGAAGATAAGGAAAAAGAAGAGGTTGCTGCTGTCAAAATCCAAGCTGCC
TTCCGGGGACACATAGCCAGAGAGGAGGCAAAGAAAATGAAAACAAATAGTCTTCA
AAATGAGGAAAAAGAGGAAAACAAGTGAGGACACTGGTTTTACCTCCAGGAAACATG
AAAAATAATCCAAATCCATCAACCTTCTTATTAATGTCATTTCTTCCTGAGGAAGGAA
GATTTGATGT

LRRC35 exon sequences with primers for sequencing highlighted.

Key

PRIMER
CODING EXON
START/STOP CODON
5'/3' UTR
INTRON
SNP

Exon 1
CACCACAGTATATGTTTCTGGGGTCAGTAACTAAGTATCATATGTCTTGC
AATGCCTAGTCTAGTACTTTATACATAATGGGTTTTCAATAAATACTTCT
TATGTGAGTAAATACATAATTACATAATTTGATTATTTCTTGGTTTCTTG
TAG**CATTTTAAGAAAGAAAG**ATGGATCAACCTAGTGGAAGAAGTTTCATG
CAAGTATTATGTGAAAAATATAGTCCTGAAAATTTTCCTTATCGCCGTGG
CCCGGGGATGGGAGTCCATGTCCCAGCCACACCTCAGGGCTCTCCTATGA
AAGGTAAGAAAGATGGGACCTAAAACACTTATTTAGTGGAGCTTACTTAG
GATGTTAAATCACAGTGTTGCCTTGTTTGCAAATGATTTATTTTATTTTT

Exon 2
CAGTACTTCTGAACTCCTCCTGGGCAATAACTGGCTGTCTCCTCTTAGGA
TTCAGTACCCTTTGGTCCCATTATATCCTGTATTTTCTAGATCCTTATGT
AATTTCACACAGTTTGATTTTTGTTTCTTGACACTTCCTGGGGGCTGAAC
ATATGTAGACAAGAATTTGGCTGATATAGAAAACATTTTGTGTCTCTCAT
CATTCAGATCGCCTCAACCTCCCAAGTGTACTAG TGTTGAACAGCTGTGG
AATAACCTGTGCAGGAGATGAAAAAGAAATTGCTGCTTTCTGCGCTCATG
TGTCGGAACTAGATCTTTCTGACAACAAACTCGAAGACTGGCATGAGGTG
AAGTTTTTATATTGCTACATTTTAGTGAAAAGCAGCAATAACCATTGTGT
CATTGTTTATGTTATTATATGTTCTGCTAAATTCTGTTCTCACTTGATGT

Exon 3
TTTGCCAAGGTTTGTTGAAATCACTTATGTCCCTCTGTGAGTACCGTCAC
TAAATAATAGGACTCCTTGATGTAAATGGCCCTTTGCCTGGAGACCAGCT
CCTCTAGGCTTTCTTGAGCATGATTGCTATTTAACAGCTCTGATTACTGC
CTGATAATGCTTTTCTTTTTTTCTCCTTAGGTCAGTAAAATTGTGTCAAA
TGTTCCTCAGTTGGAGTTTCTAAACCTGAGTTCCAACCCTCTGAATTTGT
CGGTTTTAGAAAGAACATGTGCTGGGTCCTTCTCTGGGGTTCGCAAACTT
GTCCTCAACAACAGCAAAGCTTCTTGGGAGACGGTCCACATGATACTACA
GGAGTTACCAGAGTAAGCCCAGAGAGCATGAGGGCGAGGGAGTTATGTTG
CCATCTGTGTTAGTACATAGGAGTACTGCTGATCTCCCACGCAAGAAATA
CTAGAACTGCAGATTGAGTGGAGAAAGAGGTTAGAAAAACGAAGTTATTG

Exon 4
AGTAATTTCGCACTACCTTATATCATTCTTCGCTATTTCAGTGTGTACTT
CAGGTTTTGTCTTCCCAGCTATATTTTGAACTTCAGGAAGACATCTTGGT
ATTTTCTCCACAATTCCTAGCAGAGTCTCACATATAGTAGGCACTTACTT
ATTTATTGAATGAATGAAAATAGGTGTAAAGTAGAAATTAAACTGCTTTA
AACAATGAAATATTTCCTTTTTTCTCTGCAGTTTGGAGGAGCTCTTCCTG
TGCCTTAATGACTATGAAACAGTGTCTTGTCCTTCTATTTGCTGTCATTC
TCTTAAGCTACTACATATAACAGACAATAACCTCCAAGACTGGACTGAAA
TACGAAAGTTAGGAGTTATGTTTCCTTCACTGGATACCCTCGTCCTGGCC
AACAATCATTTGAATGCTATTGAGGAGCCTGATGATTCATTGGCCAGGTT
GTTTCCTAATCTTCGATCCATCAGCCTCCACAAGTCAGGTGAGGTTCAGG
CTTGTTCTTATTCTACATGCAAATTAACCTGAGCATATGCATGAAATACA
ATGAAAGTATATTTCTTTCAGTCACAGTTTTACCTTTTTTAGAGTGAATT
TTCTATATGACATGCAGATGCCTATGATTGGCTAAATTTAAGAGATTCAT

Exon 5

AAGTCACTTTGGTAGAAAATGAGTAGATGAGAGGATGTTAGACAATAACA
CTCTCTTAACCCGTGGATCACAGAGGTAGGAGAGGCTATGGGGAAGTTTT
TGTGGTTTTAACTGTGTTCTGCTCTTGCCTTGGATGAAGAAAACTTTCAG
ACCGTTTAGTTTATGATCATTTTAACATTTCAGAAGTTACTAATTGAGCT
AATATTTTTGCTATGTTTCTCTTCTTATTTATGTGCATCTCTGGATTTAC
TTAAACAATATGTTCTCAAATTAGGTTTGCAGTCCTGGGAAGACATTGAT
AAACTAAATTCATTTCCCAAACTGGAAGAAGTGAGATTGTTAGGAATTCC
TCTTCTGCAGCCATATACCACCGAGGAGCGAAGGAAATTGGTAATAGCCA
GGTCTGTTGTCCTGATCGTTTTGCTTTATTTTTGTGAGGCCTTATTGTTT
CATTAAAGGAATAATGCACTGTTTAGTGGGAGTGCACTATGAAATTATTC

Exon 6

GTTCATGTCAGTAGTTGTATACAAATGGAACTACATCAAATGTGTCCTTT
CAAGAAAATAAGACTGGGCCACAAGCCAATTAGTTTCTTTCTAAAGCTAA
ATATATTAGTATCTTAAAAAATGTCTTTATTTTGGTTTTAACTTATAGAT
TGCCATCAGTTTCCAAACTTAATGGCAGCGTTGTTACTGATGGTGAACGA
GAAGATTCTGAGAGATTTTTTATTCGTTACTATGTGGATGTTCCACAGGA
AGAAGTGCCATTCAGGTAAAAAATTCCCCATTGGCCAAACACTTTAGAGG
GTGGTGATGCCAGTTAAGAACTCTAGTGTTAGAGCAAAATCTAATCATTT
GTTATTGGACCCTTCTTAGAGAACTTTGTTAATACGTTAAAAAAAAGGAA
GAAACACAAGAGCAAACCTATTTTATCTCATTTTCCACAGTATATTTCTT

Exon 7

ACTCATATTAGATCACAGAGTCCATAGGAGTCATGTGATAGAATAAGATT
TTTTAAAGATGTTGAGCCTAGTAATATTGTTTATCTGTAGTTGGGAAGAA
GTCCACAGTACATGTGCTAGTAGTTTAAGCTGACAGTGTTGTTTTTAATC
CTTCTAGGTATCATGAACTGATCACTAAATATGGGAAGTTGGAGCCTTTG
GCAGAAGTGGACCTAAGACCCCAGAGCAGTGCAAAAGTAGAAGTCCACTT
TAACGATCAGGTGGAAGAAATGAGCATTCGTCTGGACCAAACAGTGGCAG
AACTAAAGAAACAGTTAAAAACTCTAGTACAATTACCCACAAGCAACATG
CTTCTCTACTATTTTGACCATGAAGCACCCTTTGGCCCAGAGGAAATGAA
GTACAGCTCTCGGGCATTGCATTCCTTTGGCATTAGGGATGGAGATAAAA
TTTACGTGGAATCCAAAACAAAATAA**CCTCTACCAGCCTTGTGAAAAACA**
TACACATAAGGACTTGTTGCAGGGCATTTGTTTTTAATGTGGTTTTCTTT

FLJ35808 exon sequences with primers for sequencing highlighted

Key

PRIMER
CODING EXON
START/STOP CODON
5'/3' UTR
INTRON
SNP

Exon 1
GCAACACAGACAGTGTGCGTGGTTACTGTCATTTTTCCTTAGTCTAAGGGCACAGAA
TGGAAAGGTTCTCATTTTCCTCCTCCAATGAGAATTGGCACCCCAACAGCAGGCCCT
GGGCCCTCGGCTGTGATGGAGGGGTGTCTCGGGGTAGCAGAACTAAGGTGAGGTTG
CAAGGCTTCAAGTGTTGTTCTGCCTCTTCCATTTCCCCTGTCCTCTTCCTCTGTTCAT
CTCCCTCTCACC

Exon 2
CGTTCATCCTCGCTGGCTCTGGAGCCTTGTCTCCGCCAGGAGTTTCTGGTTTCCAGC
ATGCCATGGCCTCGGGTGACCAGAAGCCAGCTACACTGCCTCCTGGCCTGTGGTTCC
CTTCCAGGAAGCTCTCCACCTTCAGTGCATACTTGGAGGACCACAGCTACAACGTGG
AGCAGATATGGAGGGATATTGAGGACGTCATCATCAAGACCCTCATCTCGGCCCACC
CCATCATCAGGCATAACTACCACACCTGCTTCCCCAACCACACACTCAACAGCGCCT
GCTTTGAGATCCTGGGCTTTGACATTTTGTTGGACCACAAACTCAAACCCTGGCTGC
TGGAGGTAGGGAAGTTTGGGCGAGGGGCGAAAACGAACTCCTGTTATTGGGGCCAG
CCCGCTTCAAGGATTCCGCCCTCCTGGCTCACCTAAATTGTTTCAACTTCCCAGGCC
ACTTCCTCCCATGGTGCCTGCTGCC

Exon 3
CTTGGTAGTGAGGGCCCTGAGGTGGGCTTTGGTAGTGCCCTGGAGGTGGCACCTCG
GCCTCACTCTCTGCTTTCATTCTGGACGGCTTTGTCACCAGCTGCCAATCTTGCCCTC
ACAGGTCAACCACTCTCCAAGCTTCTCCACCGACTCTCGGTTGGATAAAGAGGTGAA
AGATGGTCTGCTGTATGACACCTTAGTCCTGATCAACCTGGAAAGCTGTGACAAGAA
GAAAGTCTTGGAGGAGGAGAGACAACGGGGGCAGTTCCTGCAGCAGTGTTGTTCTC
GGGAGATGAGGTACATTCGCACTCCCCAGCCTTGCGGGGCTCTCTCTCTGACTCCCC
TGCTCCCAATAAGAATGAT

Exon 4
CTGATTGACCGCCATTATGTGCAAGGCCTCCATCTGGCCCTAGACGACCTCCAGCTT
GGACATATCCCTTCCAGGCTTTGTCCCAAGAAAACCTGACAGCAGCCCCATGTTCTG
CTTCCCAACAGGATTGAGGAAGCCAAGGGTTTCCGGGCCGTGCAGTTAAAGAAAAC
TGAAACGTATGAGAAGGAAAACTGTGGAGGGTTCCGACTGATTTATCCCAGTCTGAA
TTCGGAGAAGTATGAGAAGTTTTTCCAGGACAACAACTCCCTCTTCCAGAATACTGT
TGCTTCCAGGGCTCGGGAGGAGTATGCCCGGTAGGAAGACATCCGGGATGAGGTCG
GTGGAGGGGTTCTGGGGATTGTGTGTGATCTATTGGCTGACTGGCCTGGAGTGGCA
GGGAGAGTCCCCAGCCATGAAAGT

Exon 5
TGTGGAACTTGCTAACTTTTCGCCCAGTCATGTCCATCCCTCAAGGAGAATGCCTGC
CTGCGCAGCGCATTTCTAACCCCCATGATGTTCACTTGTTCTTCATGGGCAGGCAAC
TGATCCAGGAGCTGAGACTAAAACGGGAGAAAAAGCCCTTCCAAATGAAGAAGAAG
GTAGAGATGCAGGGGGAATCGGCAGGCGAGCAAGTGAGAAAGAAGGGCATGAGGG
GCTGGCAACAGAAACAACAGCAGAAAGACAAGGCCGCCACCCAAGCCTCCAAACAG
GTAAACCTGGATTGAGGGGAACACACGTAGCCACACCCTTCCCCACCTGGGCCTGAT
CCGTGGAGGGGTGCTGAGGAGACCACTGCTCAGAGCCCAGACACACCGAAAGGCTG
TGCAGAGGGCTGGGCAACGACGGGGC

Exon 6

CTTCTCCTCATCTCTTTTTCACCTACCCCTGGGTGCAGATGAATCTGGAAGTGGATA
GAGCTGGGTTCCCATGGCTGTGTGTTGTGGTTTCTTCCTCCCTCAGTACATCCAGCC
ATTGACATTAGTATCCTACACACCTGACTTGCTCTTGAGTGTCAGAGGTGAAAGGAA
AAATGAAACAGACAGCAGCCTCAACCAGGAGGCTCCCACGGAGGAGGCCAGCTCTG
TTTTCCCCAAGCTGACGTCTGCGAAGCCCTTCAGTTCTCTACCCGATCTGAGGAATA
TCAATCTCAGCAGCTCGAAGTTGGAGCCCAGTAAACCCAACTTCAGCATCAAGGAGG
CCAAGTCTGCCTCTGCAGTGAACGTATTCACTGGCACTGTGGTAAGTAGTGAGCCGA
GCTCTGGGAGGGAGTGGTGGGAGTCTTGCTTACTGGTCCTTTCTCTACCCCCAACTT
GCTGTGACCCTTTGAGTAAGCCACT

Exon 7

TCTCAAAAAATAAAAATAAAAAAAAAAATAGCATAGGAAGGAACCCTTTGCTTTCCAC
AGTCTCATTTTGAACTGTCTTTTCTTCCAGCACTTAACCTCCGTAGAAACCACCCCAG
AATCCACCACCCAACTCTCAATCTCCCCAAAGTCTCCGCCAACCCTGGCTGTGACCG
CCAGCTCTGACTACAGTGGCCCAGAGACGGACAGGGTGGTATCCTTTAAATGCAAG
AAGCAGCAGACCCCTCCACACTTAACCCAGAAGAAAATGTTAAAATCTTTTCTGCCC
ACAAAATCCAAGAGCTTCTGGGAGAGTCCGAACACAAACTGGACTTTGCTAAAGAGT
GACATGAACAAGCCACATTTGATATCCGAGCTACTCACCAAGCTTCAACTGAGTGGG
AAGCTCTCCTTCTTCCCAGCTCACTACAACCCCAAGCTGGGGATGAATAACCTGTCA
CGTGAGTGCCAGTGTGTACAGGGATGTGCCAGGAGCTTAAAGGGGAGAGGGACCGA
TAGAACCAGTCCTAATCTTTGGGGAGCT

Exon 8

ATGGGACCCTACTGGGAGGTGGTCTTCACCAGGAATTTGTGCTGAATCCAGCAGCAC
AAATCGCTGATGCTTCTGACTTTGCCTTTCCAGAAAACCCCTCCCTGCCTGGGGAGT
GCCACTCCCGCAGTGACAGCTCTGGCGAGAAGAGGCAGCTGGATGTGTCCTCCCTC
CTCTTGCAGAGTCCTCAGAGCTATAATGTTACTCTGAGGGACCTGCTGGTGATTGCC
ACTCCAGCCCAACTGGATCCAAGGCCTTGTAGAAGCCACGCAAGTGCTATGAGGGA
CCCATGTATGCAGGATCAAGAAGCATACAGCCATTGCCTGATCTCTGGCCAAAAAGG
ATGTGAGAGGAGCTAGGTAGGCCCACACATATTCCCTTAATGAATTTGGTGCGTTGG
CAGGTGCTCTGTGTTTTTCTGGA

LRRC46 exon sequences with primers for sequencing highlighted.

Key

PRIMER
CODING EXON
START/STOP CODON
5'/3' UTR
INTRON

Exon 1 and 2

AGTCTCTGAGTTTCTCTCTCCTCCTGCCATCCTGAGTTCTGCCATCCTTAGGGGCCG
CCAAGACCTCTCTTTTCGTTCCTCTCCCGCCTCAGACCAGCAGCCTTTTATTTTAGAT
CATGTCTGGAGGTGAGTAGGGAGGTGGGACGGTGGGTAGAGCAGTTGGAAAACAGC
GGGGTAAGGCCTCCAAGAGTGAGGAAGTGTCACAGTTGGACTCATCTTTTCACAGG
GAAGTCAGCCCAGGGTCCAGAGGAAGGGGGCGTCTGCATCACTGAAGCCCTTATCA
CTAAGCGGAACTTGACTTTCCCTGAAGATGGGGAACTGTCAGAGAAGATGTGAGTG
CATGGGGGAGAAAAGGGGTGCTAGAAGCCTGAAACACTCAGGATGACCAGGTCCCT
GACGTGTACTGATTTGTCAACTGAT

Exon 3

GAAACCCCTTCCCTCCTCCGTCTCCCATCCTCCCTGCTAAGAGACCACATGAGTGGT
GCTCTAACACCACCCTTACTGACTCTTTTCCCAGGTTTCACACTCTTGATGAACTGCA
GACTGTCCGCCTGGACCGGGAGGGGATTACTACTATCAGGAACTTAGAAGGCCTCC
AGAATCTTCACAGTCTCTATCTGCAAGGGGTAACTTCTTTCTCCACCCTTCCCCTCCC
CTTGACCCCTGTGGACCTAATCTGTCTCATCTGAAAGGAGCCAACCAGGTCATCCTG
CCCATTCCCTGGTG

Exon 4

TCCCAAAGTGCCAGATCTCCAGCTGCAGAGTAGACATGGTCTTGGGGCTTTGTCCCC
GTAAAGGGCCCCTGACGCATCTGATCTCAGCTTGGTTTCCCCACTTCGGTTTCTTCT
TGCAGAATAAGATCCAGCAAATTGAGAACCTGGCTTGCATCCCCTCCTTGCGGTATG
TGGTGCCAGGGCTCAGGCAGGGGAAGAGGGGTTGGGGGAACCTAGCCATATCCCAA
AGAACTGGTTCAGATTTCTCCAAGCCTGGGTCCTTGTCAGTCAAGGACCAGCAAGCC
TAGATCTACTCCCTCA

Exon 5

CCCCAGAAGGCCATGTGTCCCCACTGGTCATCAGCTGCCTAGCCCAGGGGCCCCGG
TGTCAGGTGCCCTCCTCTTTTCTTCCAGCTGTTCCATGATTCAGCTCTGCCTCTGTAC
CCCTTCCTTCCCCTACAGCTTCCTGTCTCTGGCAGGAAACCAAATCAGGCAGGTGGA
AAACCTCCTCGACCTCCCATGCCTCCAGTTTCTGGACCTTTCTGAGAACCTGATAGA
AACATTGAAGCTGGGTAGGAACCCACTCGCCCTGGCTCACCAAACACATCCCCTGTG
GGGCCTGCCCTCCCCTCTGCACCCATCTGCCCATGCTTCTATTAGCCATGCCCCTGC
AGGTCCACATTTCATCCT

Exon 6 and 7

AACCCCATTCCCATCGCTTCATGGTCTTTTGGCTAAGATCAAGTGAGAACCCCATTT
CCTCTCTCTTTGCTGTGTGCAGATGAGTTCCCCCAGAGCCTTCTCATCCTCAACCTGT
CTGGAAACAGCTGCACCAACCAGGATGGCTACCGGTAAGGAGTGGAGGGTGGGAAG
AAAGGTCATGGCTGAGCTCTACCTGCTAACTCAGCCCAGACCCCTCTCAGCCTGCAA
ACCTGGGGTCCTGTCCATGACACCTTCTCCCCCACCTCCTACCTAGCTCATGAGCCA
CCACCCCTCCGGGTGTGAGTGCTGTGGTGCGTGTCTTTGCATCCTCCACGCCAGGGT
GCCCTCCTGGGTAACCTCTGCTCCTTCTGCTCTGCAGCGAGCTGGTGACAGAAGCCC
TGCCACTTCTCCTGGACCTGGACGGGCAGCCTGTGGTGGAGCGCTGGATTTCGGAT
GAGGAGGATGAAGCCTCAAGCGATGAGGAGTTCCCAGAGCTGAGTGGCCCATTCTG
CTCAGAACGAGGTGACCCTGCTTTCCAAGGTTTTCAGCCTCCCCAGAGCCCACCTCT
GCCCTGTGGGACATGAGGCAAGCTAAGGTGGC

Exon 8

CAGTCCTGGGCCCCAGCCTCAGGCTCCTTCCCACAAGGGACAAGGGGCCAGCCAGA
GACTTCCCTGAGCCCAGGGACCCTCCTGCTGAGCCCAGGGCTCCACCCACCCTGTAC
CCTCCACCCCCGGCCCCTCCCAGGCTTCCTCAAGGAGCTGGAGCAGGAGCTGAGCA
GGCACAGGGAGCACCGGCAACAGACGGCCCTGACAGAGCACCTGCTGAGGATGGAG
ATGCAGCCCACCCTCACCGACCTGCCCCTGCTACCTGGGGTGCCCATGGCTGGGGA
CAGCAGCCCTTCTGCCACTCCTGCGCAAGGGGAGGAGACAGTCCCTGAGGCCGTCT
CCTCACCCCAGGCCTCCTCTCCCACCAAGAAACCATGCAGTCTGATTCCCAGGGGCC
ACCAAAGCTCTTTCTGGGGAAGGAAGGGGGCACGAGCAGCCACAGCCCCCAAGGCC
TCTGTGGCTGAGGCCCCCAGCACAACCAAAACTACGGCCAAGAGAAGCAAGAAATG
ATTCTCTGTCAACCTTTCTCTACTAGTGGAGAGGAGTGGGGCCTGCCCCTCTTCTCA
GACCTCTGACCTGTGACAGAAGCCCATCCCCAGTAAAGTGTCTCTAGGCCCTGAGTA
TGCTTTTCATGTCACTTGGGGTATCTCAGGGGAAGAAACC

APPENDIX 2

Multipoint MERLIN linkage analysis for Arabic family 152 following 600 kb *Illumina* SNP screen

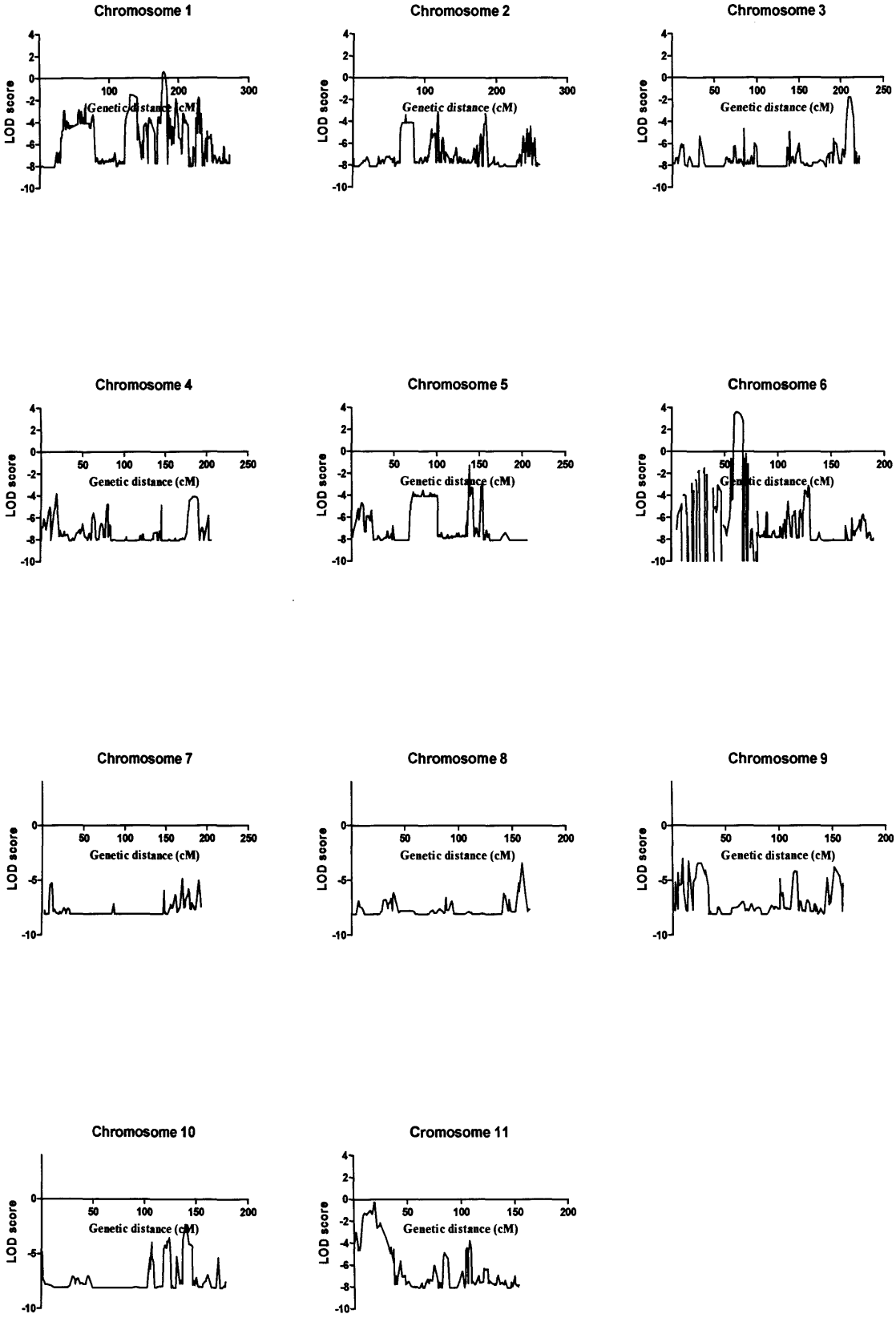
Novel in-house markers used for genotyping on chromosome 6

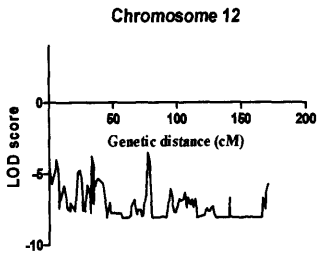
Comparative BLAST and function analysis of genes within the region of interest on chromosome 17

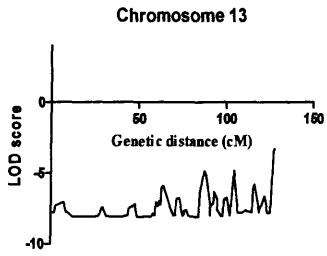
Primer sequences used for sequencing candidate genes

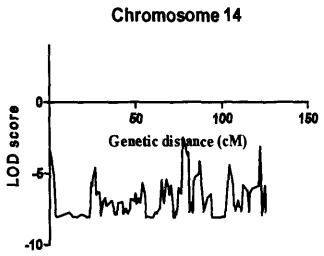
Candidate gene sequences with primers highlighted

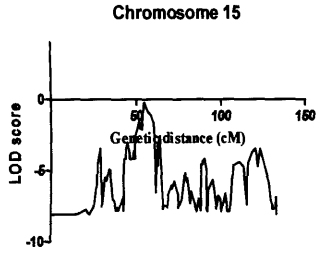
Multipoint MERLIN linkage analysis for Arabic family 152 following 600 kb *Illumina* SNP screen. LOD scores >3 on chromosome 6

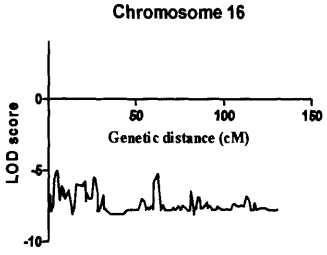


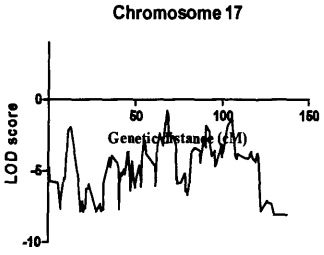

Chromosome 12
LOD score
0
-5
-10
50
100
150
200
Genetic distance (cM)

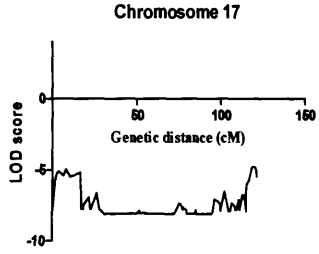

Chromosome 13
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)

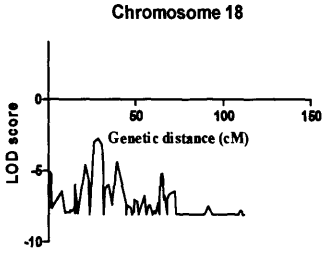

Chromosome 14
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)

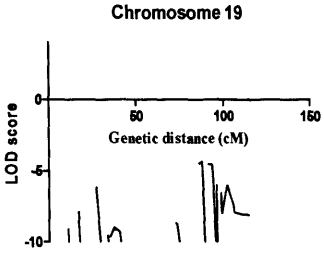

Chromosome 15
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)

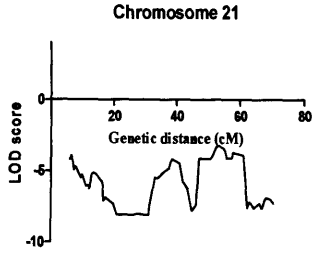

Chromosome 16
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)

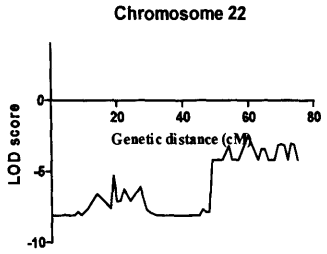

Chromosome 17
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)


Chromosome 17
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)


Chromosome 18
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)


Chromosome 19
LOD score
0
-5
-10
50
100
150
Genetic distance (cM)


Chromosome 21
LOD score
0
-5
-10
20
40
60
80
Genetic distance (cM)


Chromosome 22
LOD score
0
-5
-10
20
40
60
80
Genetic distance (cM)

Marker	Forward	Reverse
TEX27	gggtgcctaatgtaggttgg	actggtgatgagccacagg
DNAH8 in 35	tgtcaggtttagtgacctgtcc	gttctttcccacgaagctacc
TBCC	gagtcacagagggagatcctg	caagggctccttggaacag
KNSL8	ggactatagccctcagggtgac	cagggtgagaggagagacagg
UAEPal1	gctccagcttcagtgttttca	tgtggtggctcattcctgta
UAEPal4	ggctttctccatgtctatacgg	ttacaggcatgagccactgt
UAUPal2	tttgtgatctcttggcatgg	gatggcctgagaaagtctcg
UAEPal5	ctggaaattgggcaggact	gcatcttccagctaacctcct
UAEPal3	cctcccttcttttcctccag	gggctgtcagcaacttgttt
UAEPal6	gaggcatgacttgcttgaga	tcaggcagaaggagaaggaa

Novel in-house markers used for genotyping on chromosome 6

Comparative BLAST and function analysis of genes within the region of interest on chromosome 17, as described in Chapter 4, Section 4.1.9

Accession	Gene	ARAB	CERV	POMB	CHLY	PAZO	DEFLA	TRYP	CION	TETR	McKE	ELEG	fox1	Function	Domains identified by NCBI BLASTp	UniGene expression
NM_138572.1	taube nuss homolog (mouse) (TBN),	1	0	1	0	0	0	0	1	0	0	1	0	regulation of transcription	Bromodomain transcription factors	inc. eye, kidney, testis, lung NT
NM_000409.2	guanylate cyclase activator 1A (retina) (GUCA1A),	1	1	1	1	1	0	1	1	1	1	1	1	Ca2+-binding, stimulates cGMP synthesis in photorec	EF-hand, calcium binding motif	eye, testis, lung NT
NM_018415.1	transcriptional regulating factor 1 (TRERF1), transcript variant 3,	0	0	0	0	0	0	0	1	0	0	1	0	zinc-finger transcriptional regulating protein	DNA-binding domains	inc. eye, kidney, testis, lung NT
NM_015255.1	ubiquitin protein ligase E3 component n-recognin 2 (UBR2),	1	1	1	0	0	0	1	1	1	0	1	0	invd in Proteolysis	Putative zinc finger	inc. eye, kidney, testis, lung, trachea
NM_000322.2	retinal degeneration, slow (RDS),	0	0	0	0	0	0	0	0	0	0	1	0	cell surface glycoprotein found in the outer segment of both rod and cone photoreceptor cells. It may function as an adhesion molecule involved in stabilization and compaction of outer segment disks, invd in ADRP. But only in elegans		brain, eye, lung, trachea
NM_003192.1	tubulin-specific chaperone c (TBCC),	0	0	0	1	0	0	1	1	1	1	1	0	Tubulin co factor, interects with RP protein, Cofactor C is one of four proteins (cofactors A, D, E, and C) involved in the pathway leading to correctly folded beta-tubulin from folding intermediates	X-linked retinitis pigmentosa 2 gene product domian	inc. eye, kidney, testis, lung NT
NM_001004322.1	FLJ38717 protein (FLJ38717),	0	0	0	0	0	0	0	1	0	0	0	0		no putative domains	no info
NM_015349.1	KIAA0240 (KIAA0240),	0	0	0	0	0	0	0	1	0	0	1	0		no putative domains	inc. eye, kidney, testis, lung NT
NM_198486.2	ribosomal protein L7-like 1 (RPL7L1),	1	1	1	1	0	0	1	1	1	0	1	0	ribosomal protein		inc. eye, kidney, testis, lung, trachea
NM_001008739.1	similar to RIKEN cDNA 2310039H08 (LOC441150),	0	0	0	0	0	0	0	0	0	0	0	0	Riken = Ca binding protein receptor, invd in pancreatic ductal adenocarcinoma	no putative domains	inc brain, eye, lung NT
NM_138296.1	pre T-cell antigen receptor alpha (PTCRA),	0	0	0	0	0	0	0	0	0	0	0	0	a component of the pre-T cell receptor (pre-TCR), which regulates early T cell development	no putative domains	kidney, spleen, thymus, lung only
NM_006586.2	trinucleotide repeat containing 5 (TNRC5),	0	0	0	0	0	0	0	1	0	0	1	0		no putative domains	inc. eye, kidney, testis, lung NT
NM_018960.4	glycine N-methyltransferase (GNMT),	0	0	0	0	0	0	0	0	0	0	0	0	participates in the detoxification pathway in liver cells	coenzyme metabolism domains	inc kidney and lung
NM_000287.2	peroxisomal biogenesis factor 6 (PEX6),	1	1	1	1	1	0	1	1	1	0	1	0	invd in peroxisome folding and organisation	AAA-superfamily of ATPases	inc. eye, kidney, testis, lung, trachea

NM_006245.2	protein phosphatase 2, regulatory subunit B (B56), delta isoform (PPP2R5D), ,	1	1	1	1	0	0	0	1	1	0	1	0	Serine/threonine protein phosphatase		inc. eye, kidney, testis, lung, trachea
NM_014623.2	male-enhanced antigen (MEA),	0	0	0	0	0	0	0	0	0	0	0	0	regd by X and Y repressor/activator genes, respectively. Expressed in cytoplasm of spermatids, poss role in late stage of spermatogenesis	MEA1 domain	inc. eye, kidney, testis, lung NT
NM_057161.2	kelch domain containing 3 (KLHDC3),	1	1	1	1	1	0	1	1	1	0	1	0	invd in activation of the V(D)J recombination		inc. eye, kidney, testis, lung, trachea
NM_033112.2	chromosome 6 open reading frame 153 (C6orf153),	1	0	1	0	0	0	0	1	1	0	1	0		Domain of unknown function	inc. eye, kidney, testis, lung NT
NM_014780.3	cullin 7 (CUL7),	0	0	0	0	0	0	0	1	0	0	0	0	invd in cell cycle	Anaphase-promoting complex	inc. eye, kidney, testis, lung, trachea
NM_015950.3	mitochondrial ribosomal protein L2 (MRPL2), nuclear gene encoding mitochondrial protein,	1	1	1	1	0	0	0	1	0	0	1	0	mitochondrial ribosomal protein	ribosomal protein domains	inc. eye, kidney, testis, lung NT
NM_201522.1	kinesin-like 8 (KNSL8), transcript variant 2,	0	0	0	1	1	0	1	1	1	0	1	0	kinesin, microtubule motor activity	Tetratricopeptide repeat domain	inc. eye, kidney, testis, lung, trachea
NM_002821.3	PTK7 protein tyrosine kinase 7 (PTK7), transcript variant PTK7-1,	1	1	1	1	1	1	1	1	1	0	1	1	Receptor protein tyrosine kinases		inc. eye, kidney, testis, lung NT
NM_003131.2	serum response factor (c-fos serum response element-binding transcription factor) (SRF),	1	1	1	0	0	0	0	1	0	0	1	0	transcription factor: stimulates both cell proliferation and differentiation	MADS_SRF_like	inc. eye, kidney, testis, lung NT
NM_015089	p53-associated parkin-like cytoplasmic protein (PARC),	1	1	1	1	0	0	1	1	1	0	1	0	critical regulator in controlling p53 subcellular localization	In Between Ring fingers, Cullin	inc. eye, kidney, testis, lung, trachea
(NM_006443)	chromosome 6 open reading frame 108 (C6orf108), ,	0	0	0	0	0	0	0	1	1	0	0	0	The exact function of this gene is not known but studies in rat suggest a role in cellular proliferation and c-Myc-mediated transformation	catalyses the cleavage of the glycosidic bonds of 2`-deoxyribonucleosides	inc. eye, kidney, lung NT and not testis
(NM_032538)	tau tubulin kinase 1 (TTBK1),	1	1	1	1	1	1	1	1	1	0	1	1	Ser/Thr protein kinase phosphorylating tau, beta-tubulin, MAP2 and alpha-casein	Serine/Threonine protein kinases, catalytic domain	eye, kidney, testis, lung NT
NM_153320	solute carrier family 22 (organic anion transporter), member 7 (SLC22A7), ,	1	1	1	1	0	0	1	1	1	0	1	1	lc22a7 is a multispecific organic anion transporter of the liver and kidney	Arabinose efflux permease, Nitrate/nitrite transporter	kidney, no other relevant tissues
NM_206922	cysteine-rich protein 3 (CRIP3),	1	0	0	0	0	0	0	1	0	0	1	1		Zinc-binding domain	brain, testis, kidney, lung, NT
NM_014345	zinc finger protein 318 (ZNF318),	0	0	0	0	0	0	0	1	0	0	0	0	Zinc Finger	Zinc Finger	inc. eye, kidney, lung NT and not testis

NM_138572.1	ATP-binding cassette, sub-family C (CFTR/MRP), member 10 (ABCC10),	1	1	1	1	0	0	1	1	1	0	1	1	member of the MRP subfamily which is involved in multi-drug resistance, nhibition of NK cell-mediated lysis that is reversible using anti-CD94 and anti-class I mAbs	ABC-ATPase	brain, eye, kidney, lung NT
NM_000409.2	EGF-like-domain, multiple 9 (EGFL9), ,	1	0	0	1	0	0	1	1	1	0	1	1	EGF-like protein	calcium binding domain	brain, eye, lung NT
NM_018415.1	tight junction associated protein 1 (peripheral) (TJAP1),	0	0	0	0	0	0	0	1	0	0	0	0	present at tight junctions between epithelial cells	none detected	inc. eye, kidney, testis, lung, trachea
NM_015255.1	chromosome 6 open reading frame 154 (C6orf154),	1	0	0	1	0	0	1	1	0	1	1	1		Leucine-rich repeats, Tropomodulin: blocks head-to-tail association of tropomyosin along actin filaments	brain, eye, testis, lung NT
NM_000322.2	Yip1 domain family, member 3 (YIPF3),	1	0	0	0	0	0	0	1	0	0	1	0	involved in hematopoiesis	Yip1 domain (golgi protein involved in vesicular transport that interacts with GTPases)	inc. eye, kidney, testis, lung NT
NM_003192.1	polymerase (RNA) I polypeptide C, 30kDa (POLR1C), ,	1	1	1	1	0	0	1	1	1	0	1	0	RNA polymerase	RNA polymerases D	inc. eye, kidney, testis, lung, trachea
NM_001004322.1	exportin 5 (XPO5),	1	1	1	0	0	0	1	1	0	0	1	0	mediate the transport of proteins and other cargo between the nuclear and cytoplasmic compartments	Importin beta-related nuclear transport receptor	inc. eye, kidney, testis, lung, trachea
NM_015349.1	polymerase (DNA directed), eta (POLH),	1	1	1	1	0	0	1	1	1	0	1	0	damage-bypass replication protein	DNA polymerase	inc. eye, kidney, testis, lung, trachea
NM_198486.2	GTP binding protein 2 (GTPBP2),	1	1	1	1	0	0	1	1	1	0	1	0	GTP-binding protein	GTP-binding	inc. eye, kidney, testis, lung NT
NM_001008739.1	MAD2L1 binding protein (MAD2L1BP), ,	0	0	0	0	0	0	0	1	0	0	0	0	delays the onset of anaphase until all the kinetochores are attached to the spindle	none detected	inc. eye, kidney, testis, lung NT
NM_138296.1	chromosome 6 open reading frame 206 (C6orf206),	0	0	0	1	1	1	1	1	1	1	0	1	mitochondrial ribosomal protein S18A-like 1	none detected	inc. brain, eye, testis, lung NT
NM_006586.2	mitochondrial ribosomal protein S18A (MRPS18A), nuclear gene encoding mitochondrial protein,	0	0	0	0	0	0	0	1	0	0	1	0	28S subunit protein that belongs to the ribosomal protein S18P family	ribosome	inc. eye, kidney, testis, lung, trachea
NM_018960.4	vascular endothelial growth factor (VEGF), ,	0	0	0	0	0	0	0	0	0	0	0	0	increases vascular permeability, inducing angiogenesis, vasculogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis	Platelet-derived and vascular endothelial growth factors (PDGF, VEGF) family domain	inc. eye, kidney, testis, lung, trachea

Exon	Fwd	Rev
1 section 1	CATCATGTTAGGCCAATCACC	CATACAGTCTGACCGTGCAGTT
1 section 2	CAACCTGGAGTCCCAAGTCT	GGAAAGTCCAACGTGGAAAG

Primer sequences for genomic *TBCC* sequencing

Exon	Fwd	Rev
1	TTGCTCAGCTTCTCTGTCTCA	TAGAGTGAGGTGATCCCAAGC
2	CCTTCCACCTGTTCCTGAAT	TGATGGACCCTGAGGATGAT
3	AGGGGTCACTGAGAAGGCTA	TTCTAGATGGTGACTGTCCTGTG
4	TGGAACAAGACATGCAATGG	AAAGCCAAAGGAGGACCAAG
5	GCTTCCTGTCTCCTCATGGA	GCAGCAGATGGATGAGCTAAG
6 & 7	CCCATAGGTCTTTTGCAACC	GTGTGCCAGACTGTGATGCT
8 & 9	CTTGCCTCCCTCCATGTTT	GAGTTGGGCTAATGCAAGGA
10 & 11	CACCTTGCATCTCAGCATCT	CAAGAAACCTCGGCACTCTC
12 & 13	TTTCCTTCCAGTGGCTGATG	CTCCCAAAGTGCTGGGATTA
12 & 13	TGTGGGTCTCAATAGTAGGTTGTG	GATTACAGGCGTGAGCCACT
14	CTTAAGGCCACTCAGGAGGA	CTCCCTTCCCTCTTCCCTTA
15	ACTCTAGGAGCCAGTCTTTGG	CTAAGGGCAGCAGCAATCTT

Primer sequences for genomic *KNSL8* sequencing

Exon	Fwd	Rev
1 to 4	CCCTACGGTGATTCCTCTCA	GCGTACTGGATCACCAGGTT
Nested		GCAACCTTGCTGGGATCT
3 to 10	GGATTCCCTGGATGACCTCT	CAGAGCTCCCAGGTTTCTCA
8 to 15	GTACAGGAGTTTGGGTCTGT	TAAACTGCTCCCCTTCACT
Nested	ATGATGACCACAAGCCCATCT	AAGAGCCAGGAGCAATGG

Primer sequences for cDNA *KNSL8* sequencing

Exon	Fwd	Rev
1	CCCAAACTCTGGAAAGGATG	CCATTCCACTCATCTCAGATCC
nested	CGGAGCTTCAGGTCTCCAT	CGCCCCTTTCAATTGCTG
2	GCCCAACCCACTAGACAGAA	CCAGCCTGTTGCCATAATTT
3	AGCGTGATCTGTGTGGCTCT	CTGCTGGACTCCCAGACATA
4	GGTCCCAGTGCAGACAGAAA	GTGAACCACCAATTCCCACT
5	CCAGTGGAACCATAGCACCT	ACCAGACGGAGGTGGAGTTA

Primer sequences for genomic *C6ORF206* sequencing

Exon	Fwd	Rev
1 to 2	CTTCTCCTAGCAACTCGAC	TCACCTTCTGCAGCTCAGTG
2 to 5	CCTGAACTGCACAGAGTGGA	GCTTAAGAAAATCATGTTTAGACTCTG

Primer sequences for cDNA *C6ORF206* sequencing

TBCC genomic sequence with primers for sequencing highlighted

Key

Primer
5'/3' UTR
Coding exon
SNP
Upstream DNA

GATCTGACATCATCATGTTAGGCCAATCACCGGACGCGTTGGTGGGAGGCCTCACGGACA
GCGCGCCCGGAGGAAGGAAGACAAGAGAGAGGAAGCTTGAAGCCAATATGGAGTCCGT
CGTTGCTCCGCTGCTGCTGTCAGGACCGGAGACATGGAGTCCCAGCGGGACCTGAGCCTG
GTGCCTGAGCGGCTTCAGAGACGCGAACAAGAACGGCAGCTGGAAGTTGAAAGGCGGAA
ACAAAAGCGGCAGAACCAGGAGGTAGAGAAGGAGAACAGCCACTTTTTCGTCGCCACCT
TTGTTCGGGAGCGAGCGGCCGTGGAAGAGCTTCTGGAGCGCGCGGAGTCGGTCGAGCGG
CTGGAGGAGGCGGCCTCTCGGCTCCAGGGGCTGCAGAAACTAATCAACGACTCAGTTTTT
TTCCTAGCCGCTTACGACCTGCGGCAGGGACAAGAGGCGCTGGCGCGGCTGCAGGCGGCC
TTGGCCGAGCGGCGCCGGGGGCTGCAGCCCAAGAAGCGTTTCGCTTTCAAGACCCGGGGA
AAGGATGCTGCTTCGTCTACCAAAGTAGACGCGGCTCCTGGCATCCCCCCGGCAGTTGAA
AGCATACAGGACTCCCCGCTGCCCAAGAAGGCGGAAGGAGACCTCGGCCCCAGCTGGGT
CTGCGGTTTCTCCAACCTGGAGTCCCAAGTCTTGGAGAAGAGAGCCAGCGAGTTGCACCA
GCGCGACGTTCTTTTGACCGAACTGAGCAACTGCACGGTCAGACTGTATGGAAATCCCAA
CACCCTGCGGCTAACCAAGGCCCACAGCTGCAAGCTGCTCTGCGGTCCGGTGTCTACCTCT
GTTTTCCTGGAGGACTGCAGTGACTGCGTGCTGGCAGTGGCCTGCCAACAGCTCCGCATA
CACAGTACGAAAGACACCCGCATCTTCCTGCAGGTGACCAGCAGGGCCATCGTGGAGGAC
TGCAGTGGGATCCAGTTCGCCCCTTACACCTGGAGCTACCCGGAGATCGACAAGGACTTC
GAGAGCTCTGGTTTAGATAGGAGCAAAAATAACTGGAACGATGTTGACGATTTTAACTGG
CTGGCCCGGGATATGGCCTCCCCAAACTGGAGTATTCTTCCTGAAGAGGAGCGAAATATC
CAGTGGGACTAAGCAGTTGTCACTCTGTTCTTCACTCCTACCAAATACTTTCCACGTTG
GAC
TCCCCCTTATTGGGTCTCGAAGTTTACTTATTGTCACACTGTGTATGTTTTCAGCATTTT
AAGGCTAGAGATTGTAATGGGCTCCTACTTGTAATTTCCATTAAATTCGTAACAGGTATA
ACACTAAAGCATTTTTGCTATTTTCGTCATGCCTTTGAGACTGAGTCTTACTCCGTCCCC
CAGCGTGGTGGCGCGCTGGGATTACAGGCGCGCGCCACCACGCGAACTTGTATTTTTAGT
AGAGACGGGGTTTCGCCATGTTGTCCGGGCTGCTCTCGAACTCCTGACCTCAGGTGATCC
ACCCGCTTCAGCTTCCCAAAGTGCTGGCATTACAGGCGTGAGCCACCACGCCAGGGCTTT
ATTTATTTATTTTTACCACAATAGTTTGAAGCAGTAAGGGGGAAGGAGGGTGATTATATT
GCTTTGTAATGGTTTGTGATACTTGAAACATCACGGTGCATAATAAAGTAGGTCTGCAGT
AAATACATGGGTTTGTTTGGTGTGTGTGTGTGTGTGTGTGTTTTGTTTTTAAAGAGATGG
GCTCTCCTTATGTTGCCCAGGCTGGTTTCGAACTCCTGACCTCAAGCAATTCTCCTGCCT
CAGCCTCCCAAAACACTCGGATTATAGACAGTAACCAGCAGTGTCAGGGTATAGATACAT
ATTTAATAAATTCCAGAGCT

KNSL8 genomic sequence with primers highlighted

Key

Primer
Coding exon
Start/stop codon
Intron
SNP

Exon 1
CCCACCACTCTCTGAGAAAGTTGCTCAGCTTCTCTGTCTCATCTTTACACCAGCTGAA
CTCTGTTGTTTCTCCCTAGACCGGGCAAGGTCCCCCAGGCCAGGATGTCAGGCCTGG
TGTTGGGGCAGCGGGATGAGCCTGCAGGCCACCGGCTCAGCCAAGAGGAGATCCTG
GGGAGCACACGGCTGGTCAGCCAAGGGCTAGAGGCCCTACGCAGTGAACACCAGGC
CGTGCTGCAAAGCCTGTCCCAGACCATTGAGTGTCTGCAGCAGGGAGGCCATGAGG
AAGGGCTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGAAAACATTGAG
CTCGGGCTGAGTGAGGCCCAGGTGAGAGGGCAAAGGTGGTGCCAAGTGGTCCAGG
GTGGATGGAGGAGCAGTTATCTGATTAAAGGTTTGGGGCTGCTTGGGATCACCTCAC
TCTAGACGTTTGCATTCATTCATTAAATA

Exon 2
CCTTCCACCTGTTCCTGAATACTCCCACCCAGAATCTGGAGCTCAGGGGATGTGCCC
ACCTAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGTGGAGTCGGAGAAACA
GAAGCTGCGGGCTCAGGTGCGGCGGCTATGCCAGGAGAACCAGTGGCTGCGGGATG
AGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCTCAGCTG
GAGGAGGAAAAGAAGCACCTGGAGTTCCTGGGGCAGCTGCGGCAGTATGATGAGGA
TGGACATACCTCGGTGAGTGTGCACAGGCGAGACTGGCTGAGGGGTGGGCAGCCGG
GAGTTACCACAGACATGGGGGCAATTGCCCCATCATCCTCAGGGTCCATCAGCTCTA
CCCAACCTGCTCTCACCTGAAAGAAT

Exon 3
CAGAGACAGGAACAACAGTCTTCTCCTAGGTCCCATCTGTCTGTGATTTCATGGCTT
AGAGGGGTCACTGAGAAGGCTACTAGGTAGGGATATCTTGGCCCTTATATGTTGCTT
CTTCCCTTCCAGGAGGAGAAAGAAGGCGATGCCACCAAGGATTCCCTGGATGACCT
CTTTCCTAATGAGGAGGAAGAGGACCCCAGCAATGGCTGTGAGTCTGCCTCTGGAAT
GGGAGGGTGAAAAGGGGCAGCAGGAGGCTGGCCCTAGCTGTGGCCCACCCTGTACA
CAGGACAGTCACCATCTAGAAATAGGTTCTCCACTTTCTCTGGAGACAGGGCAGGGG
CTGGGCTAGTACTTCGATTT

Exon 4
CAAGTATGGAACAAGACATGCAATGGGGAGTGGAAGGGCAGACAGCAAAGGGGAGA
GGAAGAGTCCTTTGTTTATGTTCTGTGATGGTTTAGTGTCCCGTGGTCAAGGTGCTA
CAGCAGCTCAGCAGGGTGGATATGAGATCCCAGCAAGGTTGCGGACGTTGCACAAC
CTGGTGATCCAGTACGCAGCCCAAGGTCGCTATGAGGTGGCCGTGCCACTCTGTAA
GCAGGCACTAGAGGACCTGGAGCGCACATCAGGCCGTGGCCACCCTGATGTCGCCA
CCATGCTCAACATCCTTGCTTTGGTGTATCGGTGAGGACTTCCCTCCTCCAGTGCCC
AGATCTTCCCCCACATTCCCACCTTGGTCCTCCTTTGGCTTTCATTTTTCCTCTTTAC
AGGGTACCTCTCCTCCCTGTTGG

Exon 5
GGCTTCCTGTCTCCTCATGGAGGTCTCCAATCCTTAAATGTACTCAGTCCTTCCAAA
GGGAAGGAGGGAAGGAGAAGGCTTGCGGGTTCAGGGTAACTCCTGTCACTTTTGTC
TTCTTCCAGTGACCAGAATAAGTATAAGGAAGCTGCCCACCTGCTGAATGATGCCCT
TAGCATCCGGGAGAGCACCTTGGGACCTGACCATCCTGCTGTCAGTATTCCTTGCCC
TCCCCACCCCACGCCCCCGCACCCCCCACCATTGCTGTTTTGGCCTCTCTTAGCTCAT
CCATCTGCTGCCCTCA

Exon 6 and 7

CAGATTCCCATAGGTCTTTTGCAACCCAAATGTCTGATGGGGACAGGCCTGTCTTCA
TAGGTGGCTGCCACACTCAACAATTTGGCTGTGCTCTATGGCAAAAGGGGCAAGTAC
AAGGAGGCAGAGCCTCTGTGCCAGCGGGCACTGGAGATTCGAGAAAAGGTACCCAT
GCCCTCTCTCCCTTCTTCTCTTGTCGCTGTGACCCTTCTCTACATGGCTCCTCTGGTC
TAACCCTCCTCCTTCACTTTTTTTTTCTCACCCCCATCTCTGCTTTGTTCTCTCTTCCT
TCATTCCCTTTCCTCTGCTTTCCACATGTGTGTGCCTACATGCAAACACACGTGTGCT
TGTGCACACACATGTTATCATAGCCATTTATCCACTCCTTTGTTCCCTTTCAGGTCCT
GGGCACGAATCATCCAGATGTGGCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCA
AAACCAGGGCAAGTATGAGGCCGTGGAACGCTACTACCAGCGAGCACTGGCCATCT
ACGAGGGGCAGCTGGGGCCGGACAACCCTAATGTAGCCCGGACCAAGAACAACCTG
GTATGGGAGGAGGGACAAAGTAGGTGGAAGAAACGGAGAGGGGGGCACAGAGGTG
GGGGGAGGGGGGGCAGGCGGAGAGGCCACTTGGGTCCTAGAGTCAGGAAGCATCA
CAGTCTGGCACACTGCTCTCATCAACTTTGTGCTCATG

Exon 8 and 9

CTTGTCTAAAAAAAAAAAAAAAAAAAAGACCTTGCCTCCCTCCATGTTTTGTTCCTGT
ATTTTTGCCTTTCTGTCCTCCAGGCTTCCTGTTACCTGAAACAGGGCAAATATGCTGA
GGCTGAGACACTATACAAAGAGATCCTGACCCGTGCCCATGTACAGGAGTTTGGGTC
TGTGGATGGTGAGTGGTGAGGGCCTGGGGAAGGGGCTGTGGGGAAGAAAAGAAATT
CAGACCAAAGGTCGGGGGTTAGGCCACAGCGGTTCCCAGGGGGAGCCAGACCCCTT
CAGTCCAGCCTGTCAGCCTCTGGGTCCTGGCTCTCCGACACTGTCTAGCTCCCATCT
CTAACCTCCCCACCCCATGTATCACCCCTAGATGACCACAAGCCCATCTGGATGCAT
GCAGAGGAGCGGGAGGAAATGAGCAAAGTGAGTGGGGGGAAGGGGGGCCAGCCTG
GGGAGCTCAAGGCTACCGGGGTCTCTGTCTTACTCCTTGCATTAGCCCAACTCTCAC
TTCCCATCCCAGCACCAGCCCCAGATGCATGGTGTCCTCCCCAGCCCAGCCTGCACC

Exon 10 and 11

CTAGCCCCATGCCACCTTGCATCTCAGCATCTCTGTATTCTTATATTTGGCTCTCCCT
GCCCTTCTTTCTGGCCTCTCTCTGTCCTGCAGAGCCGGCACCATGAGGGTGGGACAC
CCTATGCTGAGTATGGAGGCTGGTACAAGGCCTGCAAAGTGAGCAGGTGAGCTGAC
AGTGAAAAGCCAGCTTGGCCTCCGATGCCCTCCATCCAGTCCCTGCTCCCAGCCTCC
CAGACCTTTTCCCTTCCTCCCGTAGACTTTCCCTCAGCTTTAGCTCTGTTCCCTCTCC
TATTTTCTTATTTTTTTTCCTCTTGTCTTGCTTGGAAGACAAATGTTTTGTTTTCTGAA
CTCATTCTCTCTTCCTTCTCTGGTCTCTTTCTCACCACAGCCCCACAGTGAACACTAC
TCTGAGAAACCTGGGAGCTCTGTATAGGCGCCAGGGAAAGCTGGAGGCTGCTGAGA
CCCTGGAGGAATGTGCCCTGCGGTCCCGGAGACAGGTCAGAAGCCCAGAGGGGAAG
GAGGTCCTAGAGGGAAGGGCCCTGGGGATGAGAGAGTGCCGAGGTTTCTTGGGAGT
C

Exon 12 and 13

TTAGATTGTGGGTCTCAATAGTAGGTTGTGAAATCAATTCAGTATGGGCCATGGGGA
TAAAAATATATACAGATCTATATATAGCAATGCATTGCAAATAGAAAAGGTAAGTATT
ATTCCAGTAGGTATAAACTGAGACATCACAAGGAAAAATGTTTTCCTTCCAGTGGCT
GATGGTGAAAAAAATTTGAAAGCCCCATCCTGAGGAGTTAAGGAGTAGGAAGTCCC
AGGTCCTTCCTTTTCCTTCCAGGGCACTGACCCTATCAGCCAGACGAAGGTGGCAGA
GCTGCTTGGGGAGAGTGATGGTAGAAGGACCTCCCAGGAGGGCCCTGGAGACAGTG
TGAAATTCGAGGGAGGTGAAGATGCTTCTGTGGCTGTGGAGTGGTCCGGGGTAAGT
CTGATCATTGCCTTTTCATCTCTCCTGCCCACTCCTGCTGGTGGTGGTGAGAGGCTC
TTGGCCATGCTGTCCCAGTCTAGGTAGTGTATTCCTTCAGGGGTATCTCTGGGAGCC
AGTTTTTGACAGGACTTTTATTTCCAGTCACTGGGCCTTGGGGGTGGGGGATGTGTG
CAGAATAAATGCTAGGAGTAGACTCAGAAACCCATCTGTGTGGGATCCTTGTAGCTT
TCATTGCTCACTCCTTTCTGCCAGGATGGCAGTGGGACCCTGCAGAGGAGTGGCTCT
CTTGGCAAGATCCGGGATGTGCTCCGCAGAAGCAGTGAACTCTTGGTGAGGAAGCT
CCAGGGGACTGAGCCTCGGCCCTCCAGGTATACATGGAAGTCAAGATACAGTAAAG
TGGCCGGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG

Exon 14

AGGCCCACAGCCTTAAGGCCACTCAGGAGGAAGAGAGGAGGCCTCAATTCTTCCGT
TTTCCATTGTAGCAGCAACATGAAGCGAGCAGCCTCCTTGAACTATCTGAACCAACC
TAGTGCAGCACCCCTCCAGGTGAGAGCAGTGCTTGTGAGCATATTGGTGGGTGAGG
AAAAAAGGACAGCCAGTGAGCAAGCCCAGTGGAGTAAGGGAAGAGGGAAGGGAGA
GGAGTCATCTGGGATG

Exon 15

GGACTCTAGGAGCCAGTCTTTGGGAGGCCGAGGTCCCTGAGCTGATGGGGTAGTGT
TGTCTGTTTCAGGTCTCCCGGGGCCTCAGTGCCAGCACCATGGACCTCTCTTCAAGC
AGCTGACATTCAACCCGGCCCCCAGGTCTGCTGGGTCCCCCCACCCCCACAGCCCTC
ACAGCATTCCCCATTGCTCCTGGCTCTTCCCCACCCCTAGGTGGGACAGTGAAGGGG
AGCAGTTTAACCAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAA
TCCCTCTTAGGAAGGA

KNSL8 cDNA sequence with primers highlighted

Exons alternate blue/black type

ACTCAGTGACCTCGGGGTCGTTAGGCCTCCACGTCCTCGCTGGAGCTGCGGACCTCAGTC
GTCCATCCACACCTCCCCCCTGCCGCCCCCGCCCATCTCACTCTCCCAACCCTGGGCACC
AGGGCAGGTCGACAGCCCGAGGCACTCCTCTACTGCGTTTCCAGGCCCAGGAGACCCACT
CCAGCGCCGCGGTTCCTTGCGACCCGGCAGCCGTTCAGCAGACCTCCCATTCTTTCCTGC
ATCATCGGGCATTGTGGAGGCACCCTACGGTGATTCCTCTCAGAGCCTGTTTGTAGGTGA
CCGTGACAGAGACAAAACCATGGAAGTGACTGGGTTTGGAGTGACCAGACCGGGCAAGGT
CCCCCAGGCCAGGATGTCAGGCCTGGTGTTGGGGCAGCGGGATGAGCCTGCAGGCCACCG start codon
GCTCAGCCAAGAGGAGATCCTGGGGAGCACACGGCTGGTCAGCCAAGGGCTAGAGGCCCT
ACGCAGTGAACACCAGGCCGTGCTGCAAAGCCTGTCCCAGACCATTGAGTGTCTGCAGC
GGGAGGCCATGAGGAAGGGCTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGA
AAACATTGAGCTCGGGCTGAGTGAGGCCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAG
CACAGTGGAGTCGGAGAAACAGAAGCTGCGGGCTCAGGTGCGGCGGCTATGCCAGGAGAA
CCAGTGGCTGCGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGC
TGTGGCTCAGCTGGAGGAGGAAAAGAAGCACCTGGAGTTCCTGGGGCAGCTGCGGCAGTA
TGATGAGGATGGACATACCTCGGAGGAGAAAGAAGGCGATGCCACCAAGGATTCCCTGGA
TGACCTCTTTCCTAATGAGGAGGAAGAGGACCCCAGCAATGGCTTGTCCCGTGGTCAAGG
TGCTACAGCAGCTCAGCAGGGTGGATATGAGATCCCAGCAAGGTTGCGGACGTTGCACAA
CCTGGTGATCCAGTACGCAGCCCAAGGTCGCTATGAGGTGGCCGTGCCACTCTGTAAGCA
GGCACTAGAGGACCTGGAGCGCACATCAGGCCGTGGCCACCCTGATGTCGCCACCATGCT
CAACATCCTTGCTTTGGTGTATCGTGACCAGAATAAGTATAAGGAAGCTGCCCACCTGCT
GAATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTGACCATCCTGCTGTGGCTGC
CACACTCAACAATTTGGCTGTGCTCTATGGCAAAAGGGGCAAGTACAAGGAGGCAGAGCC
TCTGTGCCAGCGGGCACTGGAGATTCGAGAAAAGGTCCTGGGCACGAATCATCCAGATGT
GGCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCAAAACCAGGGCAAGTATGAGGCCGT
GGAACGCTACTACCAGCGAGCACTGGCCATCTACGAGGGGCAGCTGGGGCCGGACAACCC
TAATGTAGCCCGGACCAAGAACAACCTGGCTTCCTGTTACCTGAAACAGGGCAAATATGC
TGAGGCTGAGACACTATACAAAGAGATCCTGACCCGTGCCCATGTACAGGAGTTTGGGTC
TGTGGATGATGACCACAAGCCCATCTGGATGCATGCAGAGGAGCGGGAGGAAATGAGCAA
AAGCCGGCACCATGAGGGTGGGACACCCTATGCTGAGTATGGAGGCTGGTACAAGGCCTG
CAAAGTGAGCAGCCCCACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCG
CCAGGGAAAGCTGGAGGCTGCTGAGACCCTGGAGGAATGTGCCCTGCGGTCCCGGAGACA
GGGCACTGACCCTATCAGCCAGACGAAGGTGGCAGAGCTGCTTGGGGAGAGTGATGGTAG
AAGGACCTCCCAGGAGGGCCCTGGAGACAGTGTGAAATTCGAGGGAGGTGAAGATGCTTC
TGTGGCTGTGGAGTGGTCCGGGGATGGCAGTGGGACCCTGCAGAGGAGTGGCTCTCTTGG
CAAGATCCGGGATGTGCTCCGCAGAAGCAGTGAACTCTTGGTGAGGAAGCTCCAGGGGAC
TGAGCCTCGGCCCTCCAGCAGCAACATGAAGCGAGCAGCCTCCTTGAACTATCTGAACCA
ACCTAGTGCAGCACCCCTCCAGGTCTCCCGGGGCCTCAGTGCCAGCACCATGGACCTCTC
TTCAAGCAGCTGAACATTCAACCCGGCCCCCAGGTCTGCTGGGTCCCCCCACCCCCACAGC stop codon
CCTCACAGCATTCCCCATTGCTCCTGGCTCTTCCCCACCCCTAGGTGGGACAGTGAAGGG
GAGCAGTTTAACCAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATC
CCTCTTAGGAAGGACCCTCAGGACACCCTCTCTGCACCCTGTGGTCCTCTAGAGTAGCTA
GCTCTGAGGCCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTGCTGC
TGGCTTTTCAGTCAGAGGGTTGGGGGCTGGCCAGCCAAGCTGCCTTGCCCTGGCCGCTCT
TACTCCCTCCCTCTGCTGTCTCACTTCAGGTCCATGTATTTCACTTTTCTTAAATAAAAG
AATCAGGTAACCTTTC

C6ORF206 genomic sequence with primers for sequencing highlighted

Key

First round PCR primer
Nested primer
Intron
Coding exon
Start/stop codon
Base positions indicated at the sides

Exon 1
43720581 CTGTAGGGGAAGCCATACTGTGGCCGTGGCAACTGAAGACATCCCGTCCCAAACTCTGGA 43720640
43720641 AAGGATGAGGGAGGGGGCTGATGCCTAGGCCTCCGTGGGGAGGAGGTCTGGGCGGTAGGC 43720700
43720701 GGTGGGGTGGGTGGAACCTAGCGTCCAGCTAGGCGGAAGGGCGGAGCTTCAGGTCTCCAT 43720760
43720761 GGAGGCGGCTTCTCCTAGCAACTCGACGGGCGTTGAGCGGAGCCGCTGACCTGATGGACG 43720820
43720821 CCGACAGCCTCCTGCTGTCTCTGGAGCTGGCGTCCGGCAGTGGGCAGGGCCTCAGCCCGG 43720880
43720881 ACCGTCGGGCCTCGCTGCTCACGTCTCTTATGCTGGTTAAGCGCGACTACCGCTATGATC 43720940
43720941 GGGTTCTCTTCTGGGGCCGCATCCTTGGCCTCGTCGCCGATTACTACATCGCGCAGGGCC 43721000
43721001 TGAGTGAGGACCAGCTCGCACCGCGCAAGACGCTCTATAGGTGAGGAGGCCCCCGGGACG 43721060
43721061 GGCTCCCCAGAGGGTGGCTACCTGGAGGCAGGGCGGGGTGGGCGGGTCGCAGCAATTGAA 43721120
43721121 AGGGGCGGGGCCCATGGGGCCAAGGGAGGTGTGGGCGGGATCTGAGATGAGTGGAATGGG 43721180

Exon 2
43725981 ATTTCAACAGCTTTTCCTGGGCCCAACCCACTAGACAGAAGGGAAGATAATAGGAATAGT 43726040
43726041 GGGTGAGAAGGGAGTTGGAATCCAGGGGTGATGTGAATATTGTTGGCAGCCTGAACTGCA 43726100
43726101 CAGAGTGGAGCCTCTTGCCCCCTGCCACAGAGGAGATGGTGGCGCAGTCGTCTGTGGTGA 43726160
43726161 AGGGCCGCTTCATGGGGGACCCATCATACGAATATGAACACACTGAGCTGCAGAAGGTGA 43726220
43726221 ATGAAGGTGAAAAAGTCTTTGAAGAAGAAATAGTGGTGAGTGAAGAGGAGAGCAGGAGGA 43726280
43726281 GGGCCTAAAAGAGAGCCTGGGCTCATCCAAAACCCAGCCTAATAGAAATTATGGCAACAG 43726340
43726341 GCTGGGCACGGTGGCTCACGCCTGTAATCTCAGCACTTTGAGAGGCTGAGGCGGGTGGAT 43726400

Exon 3
43731141 TGCCCAGCGTGATCTGTGTGGCTCTGGGGTCCACTACACTGCAGTGGTGCGGGTGAGATG 43731200
43731201 GCCGTGCAGAGGGACCTTGCGCCGGGGTGCCTGGGCACAGTGCCCTGGCAGGGGGGCTCC 43731260
43731261 TCTCCTGTCTCCTCAGGTCCAGATCAAGGAAGAGACCCGCTTGGTGTCTGTCATTGACCA 43731320
43731321 GATTGACAAGGCTGTGGCCATCATCCCCCGAGGCGCCCTCTTCAAGACCCCTTTTGGACC 43731380
43731381 CACCCATGTCAATCGGACCTTTGAAGGTGAGTTCTCTGGGGCCCCTCTCAAGGGCTGGGG 43731440
43731441 GTATCTTTTCCCAAGCACAGTAGAACATATGTCTGGGAGTCCAGCAGGGGGGCTTACACA 43731500

Exon 4
43732161 CTTCCTTTAGCTCTCTGACAGTGGTTGACAGGGTCTTTTGTAAGCCCTACCATGTGGTCC 43732220
43732221 CAGTGCAGACAGAAAGGGGGCTGGGGGGTTTTGGGCTAACCATCATTTTCCTGCCTGCCA 43732280
43732281 CCTCTTTCTAGGACTGTCCTTGTCTGAGGCCAAGAAGCTCAGCTCCTACTTCCATTTCAG 43732340
43732341 GGAGCCTGTTGAGCTAAAGAATAAGACCTTGCTTGAGAAGGCTGACCTGGACCCCTCCCT 43732400
43732401 GGATTTCATGGACTCCTTGGAGCATGACATTCCCAAAGGTAATAGTCCATTACCTGGAGG 43732460
43732461 CCATGGGATCTGTCCTCAGGCCATAGCAGTGGGCCATGCCACCTGCCATTTTCCTCTGAG 43732520
43732521 ACTGGAACCAAGGGCCTAGTGGGAATTGGTGGTTCAGCATGGGTTTTTGAGGGTGGAGCT 43732580

Exon 5
43746381 TCTCTCCCCAGTGGAACCATAGCACCTGGGCCTGGCTGCCCAAGGAGCCAGGGGCCACAG 43746440
43746441 CATGAAGGGCAGAGCTGTGGCTCCAGCAGCACCAGGCCTCACCTCCTGCCTGTCTTATCT 43746500
43746501 CAGGGTCCTGGAGCATCCAGATGGAGAGGGGCAATGCCCTGGTGGTGCTGCGCAGCCTGC 43746560
43746561 TCTGGCCGGGCCTCACCTTCTACCATGCTCCCCGCACCAAGAACTATGGCTACGTCTACG 43746620
43746621 TGGGCACTGGCGAGAAGAACATGGACTTGCCCTTCATGCTATAGAATGGGAGCCAGCCTG 43746680
43746681 GATGTTTTTAAACAGAGTCTAAACATGATTTTCTTAAGCTTCAGTGAACTTGGCCTGCCT 43746740
43746741 GTTCTGTCCTATCTTCTTAACTCCACCTGCGTCTGGTTCCA

C6ORF206 cDNA sequence with primers for sequencing highlighted

Exons alternate blue/black type

GGAGGCGGCTTCTCCTAGCAACTCGACGGGCGTTGAGCGGAGCCGCTGACCTG**ATG**GACG
CCGACAGCCTCCTGCTGTCTCTGGAGCTGGCGTCCGGCAGTGGGCAGGGCCTCAGCCCGG
ACCGTCGGGCCTCGCTGCTCACGTCTCTTATGCTGGTTAAGCGCGACTACCGCTATGATC
GGGTTCTCTTCTGGGGCCGCATCCTTGGCCTCGTCGCCGATTACTACATCGCGCAGGGCC
TGAGTGAGGACCAGCTCGCACCGCGCAAGACGCTCTATAGCCTGAACTGCACAGAGTGGA
GCCTCTTGCCCCCTGCCACAGAGGAGATGGTGGCGCAGTCGTCTGTGGTGAAGGGCCGCT
TCATGGGGGACCCATCATACGAATATGAACACACTGAGCTGCAGAAGGTGAATGAAGGT
GAAAAAGTCTTTGAAGAAGAAATAGTG
ATTGACAAGGCTGTGGCCATCATCCCCCGAGGCGCCCTCTTCAAGACCCCTTTTGGACCCA
CCCATGTCAATCGGACCTTTGAAG
GACTGTCCTTGTCTGAGGCCAAGAAGCTCAGCTCCTACTTCCATTTCAGGGAGCCTGTTGA
GCTAAAGAATAAGACCTTGCTTGAGAAGGCTGACCTGGACCCCTCCCTGGATTTCATGGA
CTCCTTGGAGCATGACATTCCCAAAG
GGTCCTGGAGCATCCAGATGGAGAGGGGCAATGCCCTGGTGGTGCTGCGCAGCCTGCTCT
GGCCGGGCCTCACCTTCTACCATGCTCCCCGCACCAAGAACTATGGCTACGTCTACGTGG
GCACTGGCGAGAAGAACATGGACTTGCCCTTCATGCTA**TAG**AATGGGAGCCAGCCTGGAT
GTTTTTAAACAGAGTCTAAACATGATTTTCTTAAGCTTC

APPENDIX 3

Mouse *RSP9* transcript with primers for *in situ* hybridization probe highlighted

Primer sequences used for the cloning of mouse *RSP9* for *in situ* hybridization probe synthesis

Zebrafish *C6ORF206/RSP9* cDNA sequence with primers highlighted for RT-PCR

Primer sequences used for RT-PCR analysis of morpholino effects on splicing

Morpholino oligonucleotide sequence

Raw data for zebrafish olfactory pit cilia analysis (wild type embryos)
Raw data for zebrafish olfactory pit cilia analysis (MOex2 morphant embryos)
Raw data for zebrafish olfactory pit cilia analysis (MOex3 morphant embryos)

Chlamydomonas reinhardtii RSP9 genomic sequence
Primers sequences for amplification of *Chlamydomonas RSP9*

Raw data for *Chlamydomonas* flagellar beat frequency (FBF) analysis
Raw data for *Chlamydomonas* Immotility Index analysis

Attached CD1: videos of wild type and morphant zebrafish olfactory pit cilia motility and wild type, *pf17* and *pf17* transformant *Chlamydomonas* flagella motility.

Mouse *RSP9* transcript with primers for *in situ* hybridization probe highlighted

Alternate blue/black exons

GGAGACAGCTTCTACCCGCCAGACTAGCTTGGAGGAGAACGGCTGGCCTGATGGACGCCG
ACAGCCTCTTGTTGTCTCTGGAGTTGGCGTCTGGCAGTGGGCAGGGACTCAGCCCTGACC
GTCGGGCCTCACTGCTCACGTCCCTTATGCTGGTGAAGCGCGATTACCGCTTCGCGAGGG
TCCTTTTCTGGGGCCGCATCCTTGGCCTTGTGGCGGATTACTACATCGCACAGGGCTTGA
GTGAGGACCAGCTCGCACCACGCAAGACGCTCTACAGCCTGAACTGCACAGAGTGGAGTC
TCTTGCCCCCTGCCACAGAGGAGATGGCCATGCAGATATCTGTGGTGAGTGGCCGTTTCA
TGGGCGACCCTTCACACGAGTATGAACACACAGAGTTGCAGAAGGTTAACGAAGGAGAGA
AGGTCTTTGATGAAGAAGTGGTGGTCCAGATCAAGGAAGAGACTCGCTTGGTGTCCATCA
TTGACCAGATTGACAAGGCTGTAGCTATCATCCCTCGAGGTGCCCTCTTCAAGACCCCTT
TTGGAGTCACCCATGTCAATCGGACCTTTGAAGGCCTGCCCCTGTCCGAGGTCAGGAAGC
TTAGCTCGTACTTCCACTTCAGGGAGGCTATTGATCTGAAGAACAAGACCTTGCTGGAGA
AGTCGGACTTGGAGCCCTCGCTGGATTTCCTGGACTCCCTGGAATATGACATCCCCAGAG
GGTCTTGGAGCATCCAGATGGAAAGGGGCAACGCACTGGTGGTGCTGCGCAGCCTGCTCT
GGCCAGGCCTCACCTTCTACCACGCTCCTCGCACCAAGAACTATGGCTACATCTACGTAG
GCACAGGAGAGAAGAACATGGACTTGCCCTTCATGCTGTAGGAAGGCAGGCTGGATATTT
ATGTGGGGTCGAAACACTCTTTCATAAAGTTCAACTGTTGGG

Primer	Fwd	Rev
1st round PCR	TGGTGAGTGGCCGTTTCAT	CCATGTTCTTCTCTCCTGTGC
Nested PCR primer		TCTCCTGTGCCTACGTAGATG

Primer sequences used for the cloning of mouse *RSP9* for *in situ* hybridization probe synthesis

Zebrafish *C6ORF206/RSP9* cDNA sequence with primers highlighted for RT-PCR

Alternative blue/black exons displayed

GCTTCAAGGGAACTTAATTCAGATCATTCAAGTATTAATCACAACTCTCTGGTAAATATA
GATTGATTAGCGAATAGATCGAGATGGACTCTGATTCTCTGCATTACTCCCTGGATTTAG
CTGCTGCTAATGGTTTGACTCTGAGCAGTGAACAGAGAGCCGCTCTTCAATCCTCTCTGC
TCATAGTCAAGAGAAACTACAAATTCAGCCGGGTTTTGTTTTGGGGAAAGATTTTGGGAA
TAAAAAGTGACTATTTTATTGCACAAGGAGTTGAAGATGATGAACTGAAGAACAGGAAGT
CTTTGTATAGTTTAAACTGCGTAGATTGGCACCTGCTCCCCCCAGCCACCGAGTCCATGA
TTGCAGATGTGGCATTAGCAGCAACGGGACGATTCACAGGCGACCCGTCCCATGAGTACG
AGCACACAGAGATCCGCACAGAGGGCGAGGGGGACGAGGCTACACATGAAGAGGTCACGG
TGAAAGTGATTGAAGCAAGTAGACTGGCTGCAATAGTGTCAAACATCGATAAAGACGTGT
CTGTAGTCCCACGTGGTGCATTCATCAAGAGTCCCAATGGCAAAGTGCAGACAAACCGAA
GCTTTGGAGGCCTTCATCCCACTGAGGCTGCAAAACTGCGCAACTACCTGCACTTCCGTG
AGCCGGTTAACCTGAGGAACAAATCCATCTTGGAAATGTCTGAACTTAATCCTGCTCTTG
ACTTCCTAGATCCTTTGAGCGAAGACATTCTCAAAGGGTCATGGAGCCTGCAGCTGGATC
GGGGCGGCACTGTGTGTGTATTACGAAGTTTACTATGGCTGGGTTTCACCTTCTTCCATG
TTCCCCAGACTCCTCAACATGGGTACATATACATGGGAGATGGCCTCATGAACCTTGACC
TGCCCTTCATGTTATAACAGCCAAAAGGACTTCAGCGACACAATCTCCATCCGGTGTTGG
TGCCACATTTAAAAGAAACCAAAAGCATCAATCTTCCCCAGTGTTAATGTGACCTGGCTT
ACTTTCACTGTAATTGACTTTAATGTTATTTTATTTTAGGAAAACATAAACATGCCACCA
TGTGTGAAAA

Fwd	Rev
AGCGAATAGATCGAGATGGAC	TGGAGATTGTGTCGCTGAAG

Primer sequences used for RT-PCR analysis of morpholino effects on splicing

Morpholino ID	Sequence
MOex2	GGTGTAAGGCTTTTACCGTGACCTC
MOex3	GCTGTAAGTATACCTCCAAAGCTTC

Morpholino oligonucleotide sequences

Fish number	% clear cilia	% dyskinetic	% Immotile	Average CBF (Hz)
1	90	0	0	59.3
2	80	0	0	62.95
3	90	9	0	39.69
4	70	7	0	35.5
5	90	9	0	42.8
6	100	0	10	38.05
7	40	0	0	36.28
8	100	0	0	34.23

Raw data for zebrafish olfactory pit cilia analysis (wild type embryos)

Fish number	% clear cilia	% dyskinetic	% Immotile	Average CBF (Hz)
1	50	100	80	42.72
2	100	70	30	42.42
3	80	50	0	44.69
4	60	66	0	41.3
5	70	85	7	47
6	100	30	10	40.3
7	90	60	7	37.8

Raw data for zebrafish olfactory pit cilia analysis (MOex2 morphant embryos)

Fish number	% clear cilia	% dyskinetic	% Immotile	Average CBF (Hz)
1	20	50	0	41.7
2	40	75	5	47.2
3	60	66	33	28.2
4	90	56	30	38.7
5	70	71	30	36.9
6	70	71	7	40.7
7	50	60	5	41.5
8	30	66	7.5	45
9	100	75	0	40.4

Raw data for zebrafish olfactory pit cilia analysis (MOex3 morphant embryos)

See attached CD1 for videos of wild type and morphant zebrafish olfactory pit cilia motility.

Chlamydomonas reinhardtii RSP9 genomic sequence

Key

Exon
Intron
5′/3′ UTR
Restriction enzyme sites (BspE1) - destroyed by pf17 G del
Restriction enzyme site (FspI) - destroyed by human 3 bp del (CGC in chlamydomonas)
Primers
Primers
Primers for mutagenic insert - CGC deletion
TATA box motif

CAAAACTCTCAGCCTCAGATGGGAGAGCGGTGGGCGGCAGCTCACTTAACGTTTCCTGTCTGAAAAACTT
GCATGAAAACTTGCTCCCTGCCGGGGGCACTCTGACTCGCGTGTCTGCCTTTGCTCCCCGTCTGCCTTTG
CCACGCCCAGGAGTGCCCCCCTGGCCACGCCTTTACCCCTACCGTATGTGCCCTAGCAGTCCTGGGGGCC
ACGGAGCTCTGCCACACACGCCTCCATGATGTATGACAGATTCCACACCTCACGGATACGCCCTTTGCGT
GAAAAGGACGGCTGCCCAATGCACCCACGTTCCGGAACCGCCATAGGAGTGTCAAGCCGCCCAGGTCACG
ACACTGCCAACATGTGTGCGCAGACCGCCACCTGAGCAGTCAGCTTGTAATACTTGTGCCCAGGGTGTTG
TTTTTCCTGCTGGTATGAGCGTGCCCTCGCGCCATGTGCACATGGGCACTTTCACACACACGCTGGCCTT
GGTTTGAGGTCTTCATAAAGCATGGGATACATGACGGATATTTTTAAAAGCTTGTATTGAAGTAGGCGCG
TGAGGAGCCAAATGCGGGGAATTGCAATCTGCTCTAAGGCCAAGTTGAGCTAAGTCGCCAACAGCGCTCG
CTTGTGCGCAAACGAAATTCCCTTGATAATAATCACGACAACCTTTTATCATTATTGCTTTCTGCATAGCT
CCTAAGCGACGGTAGTATTTATCAATTCTCGTCGAACTTGAGACGAAGAACATGGTGCAGCTGGAGCCCA
ACATCACGCTGGTGCTGAAGCATCTTGCCAGCTGTGGAGCAGTGGTCTCAGCGGAGCAACAGGTTGGACC
CAGCCGAGGCCGCGTCGCACCATCTCGACTTTCCTGACGCCTTTGTAATATGTTCCGCAGGCAGCATTGG
ACCACTCAATACCCATTAAGCGTATCGAGGCAGGGCTCCGGAGCCTGACGCTGTGGGGGCGGCTAACGAC
GCTGAACGGCAAGGTAGGGGTGTGCCCAGCACCCGACGCCCCGACGCTGCGGCTCCGTCCTCCGCCCTAG
CGCCTGTGTTGCAGCCAAGCACCTGCTCAGCTTGTCCTATAGGTAGAGGCAGTTAACTGGCATGGTCCAG
TGCTATTGGATGAGGCGTGTGCTCTCTGGCATTCGCCTTGAGTGGGCCTGAGGCAGCTCGCAACTGGCAA
CGGATGGCCCCGCATGGACCCATCGCTCCTTAACCGGCGAGCGGCGGCGTTGCCCTCCTGCAGGACTACC
TAGTAGCGGAGGGCTACAATGTGGCGTCCAGTAAAGAGGGCGCCGCGGTCTATGAGACCAAATACTTCTA
CAGTCAAGACGGCGCACGCTGGAGTGACCTGCAGCCGGTGGACAGCGAAACCGCCACGCGCTGTGCCCGC
ATTAAGGGCATGCTCAGCGGCGACCCCGCTAAGAACTACGAGCTGGAGGAGAAGGACCCCAACGCACCAG
AGCCCAGCCCGGAGGCGGAGGAGGAGGTGAAGCCGCTGGTGTTCCAGATCCCGGAGCTGGCGGTGCTGCG
CTGCCGCGTGGACGCCATCGCCACCGCCACCAGCGTCATCCCCACGGTGCGTGAGTCGGGCGCAGGGGCG
GGGTGCGAGCGAGGCAAGCCGCGGTAGGGATAGAGGTATGGGGGGGCTGCCCTGGCAGGTATTGTATAGA
GCGCAAGCGGTGGTGATCGTGATGGTGATGGTGATTGTGGTGGTTGCAGGGCAGCCGGCCAGGGGTCTGG
TTGGAGGTTTGTGGGCATGCAAAGCTGTCTCTTAATACTATATCGGGTTGAGAGACCCTGGTGTGCATCT
GGACAGCCAGCACCCTGCTGTGCGGCTGCCCCTTACTGCACGGCTACACCTCCGTACTTTGATTCTAAGC
GCAGAATCACTGACAGTACACAGGCATAAGACGGCACCTCTTTCGGCATCCAACCTTGCCTAACCAACCA
TCTGCCGCTGCCGCTGAACCCCGCTATCCTCGCAGGACTCGACAATCCTCAACGCCGCCAGCCAGGTGGT
GCCCAACCGCCTGTTCGCGGGCGCCGCCTACCCCGAGAAGCTCGAGTCCTACCAGCACCGCTTCAGCCTG
CCCGGATCTGGCGTCACGCTGTCGCAGGTAAGACAAAGGTCGGCGTGTGGCAAAGTGCTCCGATACTTTA
CATAGTAATGACTTTCACTGCTTGCCCGGCGACCTGCCTGACCGGAATGGCCGCCTCCTACTGTATTTTG
TCTGTATGCGCCGCACGCCAAACGCCTGCTCGCCCCGTCTCGCCCTTTGCCCTGCTGTTTGTCCACGCGG
CGCGCAGGACCTTCGCGGGACTTGGGCTGTGCAGTACGACGCCTTCAAGGGGGTGGCGCAGGTGCGGTCG
CTGCTGTTCCCAGGCTACTTTTTCTACTATGCCGCCAACGCAGCTGACGTGGGGCAGCCTGTACGTGGGCG
ACGGCCTTTGCTACAACGACCTCATCTTCATGCTGTAGAGCGGAGTGAGCAGGATGGAGGAGGTTCCGGG
AGAACAGAGGGGGATGGGGACAATTAGATTGCGGGATGAGGCGGACAGGCGGCTGGGTGTGGACGGATGA
GGCTTCATGAGGTTCAACTGTTGGAGGATGTGCCAAACAAAGCGCACGCGGGCTAGGCCTTGACCTGGAT
GTATGGCATATGTGTCAGTGACATGGAAAAGGAAGCAAGTGGCGACGGTGGATGTACGGCATATGTGTGT
ACGATTGTACGAGTGGCTGGCGGGCATGAGCATCTGTACCCTCAGGCGCAACGCTTGGGTGACACTCGGG
AGCTGCTAGATAGGCGCGGACTAAATGGACGGGGTACGCCCGCTATCAGCCTAGGCTTGTTCATTGCGCG
GGCAAGTGCGCGGGTGTGAATGCAAGGAGGTCTGGCATTGTAAACGTTCTCGGTTCTCACAACTGGTTCG
AAAGTTTGACTGGTGCCTGCGAGGGTGCCCCCGACTCTCTTTGCCTGCAGCTGAAGCAACTGTTACAACT
TCTTGATGCTCACGGCATCAGACGTGGTATCACATGGCCAGCATACCTAGCCCTGCCCTAGCACCCAGTA
CCCCTGCGGCCGTGCCAGCATTCACTAAATCCTCCCTAGATGGCACTGCTCCTCTACATCCTTCCCACAA
GAACTCTCGGGCGCACATGGCTGTTTGCTTTGCAGCTCTGGTCCGTCACTGCTCCACGGGGCGGCATATG
AATGCGTCGTCCCGGTTCAGCGCCTCCTGCAGTGCAGCCATGGGACATGACAAGTCAGGGCACGGTTACA
CGGTATCCTGTCTACGCCAGATAGGCAGCCAACCGCCAACTCGACCAAGCGCATTGATAGCCAGCAGGCT
TTGCAAACACGCACCTCTGCGGGCGGCCGCCACACTCCGCGGCCTGGGTACAGCGGTATGGCGTGCCGTG

Purpose of primers	Fwd	Rev
PCR for restriction digest	CGCAGCTCACTTATCTCTTCCT	AGCACACGCCTCATCCAATAG
PCR for restriction digest and for cloning into vector	AGATTCCACACCTCACGGATAC	ACCAGTCAAACTTTCGAACCAG
Inducing human mutation	CGAGCTGACGTGGGGCAGCCTGTACGTGGGCGACGGCCTGAACAACGACC	

Primer sequences used for amplification of *Chlamydomonas RSP9*

Flagella beat frequency (Hz)				
Cell no.	cw15	pf17	pf17-T	pf17-Tmut
1	24	0	42	18
2	49	0	73	24
3	45	0.1	67	23
4	36	0	83	17
5	64	0	75	20
6	52	0	40	24
7	53	0	49	18
8	47	0	52	27
9	53	0	41	39
10	55	0	48	8
11	57	0	26	16
12	46	0	46	32
13	25	0	44	51
14	39	0	64	24
15	78	0	53	13
16	47	0	47	22
17	65	0	41	14
18	52	0	62	28
19	56	0	59	24
20	26	0	55	30
21	37	0		51
22		0		
23		0		
24		0		
25		0		
26		0		

Raw data for *Chlamydomonas* flagellar beat frequency (FBF) analysis

% Immotile flagella				
Cell no.	cw15	pf17	pf17-T	pf17-Tmut
1	0	100	0	40
2	0	100	5	60
3	0	100	5	80
4	0	100	7	80
5	0	99	10	70
6	5		5	60
7	5		0	70
8	0			90
9	0			85
10	0			95

Raw data for *Chlamydomonas* Immotility Index analysis

See attached CD1 for videos of wild type, *pf17* and *pf17* transformant *Chlamydomonas* flagella motility.